THE UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR: ANTHONY ROBERT ALFRED MARSHALL

TITLE OF THESIS: JUDGEMENT IN DECISION MAKING: THE EVALUATION OF PROPOSALS FOR NEW INSTRUCTIONAL PROGRAMS IN HIGHER EDUCATION.

DEGREE FOR WHICH THESIS PRESENTED: PH.D.

YEAR THIS DEGREE GRANTED: 1976

THE UNIVERSITY OF ALBERTA

JUDGEMENT IN DECISION MAKING: THE EVALUATION OF PROPOSALS FOR NEW INSTRUCTIONAL PROGRAMS IN HIGHER EDUCATION

by

© ANTHONY ROBERT ALFRED MARSHALL

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF EDUCATIONAL ADMINISTRATION

EDMONTON, ALBERTA

FALL, 1976

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled Judgement in Decision Making: The Evaluation of Proposals for New Instructional Programs in Higher Education submitted by Anthony R. A. Marshall in partial fulfilment of the requirements for the degree of Doctor of Philosophy

ABSTRACT

The purpose of the study was to investigate the way in which selected judges integrated five simultaneously presented items of information about proposals for new instructional programs in higher education. The research problem was stated in the form of five hypotheses:

1. The judgement process can be quantified.
2. A judge uses a linear model when integrating five cues which are presented simultaneously.
3. A judge's subjective weighting of criteria will differ from the weighting determined objectively.
4. A judge will not correctly identify the configural use of pairs of cues detected by objective analysis.
5. A judge will not subjectively identify the type of stimulus scale detected by objective analysis.

A one-third fractional replication of a 3^5 factorial design was used to construct the stimuli. Consequently third- and higher-order interactions were confounded. The factors were five criteria for evaluating proposals for new instructional programs in higher education. Each factor had three levels: High, Medium, and Low. Response was required on a nine-point numeric scale with the verbal anchors

"weak," "average," and "strong." Classical ANOVA was used to test for the presence of significant experimental effects. Further, classical ANOVA provided the data for computing the relative importance, $\underline{w}^2$, of the experimental effects.

The conclusion, that the judgement process can be quantified, was supported for nine of the ten judges. Nine of the ten judges used a linear model when integrating five simultaneously presented cues. Seven of the ten judges understood the relative importance of criteria in determining their judgement. Four of the ten judges understood the way relationships among cues influenced their judgement. Five of the ten judges understood the nature of the stimulus scale used for all cues.

Analysis of the judgement process, followed by in-service training, might be expected to lead to more consistent judgements. The ability to describe the judgement process, in terms of a linear model, suggests the possibility of using an automated process to perform the integration phase.

ACKNOWLEDGEMENTS

I wish to thank Dr. J.E. Seger for his encouragement, during the performance of this study, and his assistance in improving this document. I would also like to thank Dr.E. Miklos for his comments, both on the proposal for this study and this documnet.

My appreciation is extended to the participants in this study for their cooperation in providing data, and their encouragement, through expression of interest in the results.

Finally, my thanks go to my wife, Maire, for her forbearance and encouragement, and my son Niall for making the task easier.

TABLE OF CONTENTS

LIST OF TABLES

Chapter 1

PURPOSE, OVERVIEW OF RESEARCH, AND IMPORTANCE OF THE PROBLEM

This introductory chapter contains: a statement of the purpose of the study; an overview of research in the area; a discussion of the importance of the study; and an outline of the organization of the thesis.

PURPOSE OF THE STUDY

The purpose of the study was to investigate the integrative processes, used in formulating a judgement based on a number of items of information, and the judges' insight into their own cognitive processes. The judges were selected because of their involvement in a common task, the evaluation of proposals for new instructional programs in higher education, which provided the context for the experimental task.

OVERVIEW OF RESEARCH ON HUMAN JUDGEMENT

Slovic and Lichtenstein (1971) provided an extensive review of literature on the topic of information utilization in judgement and decision making. Their review included

more than 120 references relevant to the regression approach. The research hypotheses of the present study were based on the conclusions drawn by Slovic and Lichtenstein. For these reasons--the extensiveness of the review, and its relevance to the research hypotheses--the abovementioned reference (Slovic and Lichtenstein, 1971) was used as the major source for this study.

Prior to 1960 there was relatively little research on the processes and strategies employed by humans in integrating discrete items of information into a decision, although the groundwork for such studies had been laid (Slovic and Lichtenstein, 1971:653). The availability of electronic computers facilitated research, in the area of human judgement, by simplifying the task of performing the numerous calculations required by the regression approach. Thus, the application of the regression approach began, in earnest, in 1960.

The regression approach was so named (by Slovic and Lichtenstein, 1971:654) because of its characteristic use of multiple regression, and its close relative, analysis of variance (ANOVA), to the study of the use of information by a judge. Within this broad approach, Slovic and Lichtenstein distinguished two different paradigms which they labelled the "correlational" paradigm and the "ANOVA"

paradigm[1]. In the correlational paradigm, a judge's integration of information is described by means of correlational statistics. The structural model underlying ANOVA is quite similar to that of multiple regression, both being alternative formulations of a general linear model (Slovic and Lichtenstein, 1971:660). Research into human judgement was initially based on the correlational paradigm. In the late sixties some researchers turned to using the ANOVA model, because of its potential for describing non-linear use of cues.

The possibilities inherent in the application of Baye's Theorem were introduced, to psychologists, in the early sixties. This stimulated the "Bayesian" approach to the study of human judgement. The basic tenets of this approach are that opinions should be expressed in terms of subjective or personal probabilities, and that optimal revision of such opinions, in the light of relevant new information, should be accomplished via Baye's theorem. Thus, the Bayesian approach is appropriate to the study of judgement tasks involving sequential presentation of information. In the evaluation of a program proposal, all the information is simultaneously available to the judge.

---

[1]Detailed descriptions of these paradigms are contained in Chapter 2.

Much of the psychological research has used Baye's theorem as a standard against which to compare actual behaviour and to search for systematic deviations from optimality. (Slovic and Lichtenstein, 1971:665-66).

The present study was the first to investigate the judgement process in relation to the task of evaluating proposals for new instructional programs in higher education. Consequently, the major concern was to determine how judges used several items of information in formulating an overall judgement. Since the major concern of the study was how (rather than how well) judges integrated information, the regression approach was considered more appropriate than the Bayesian approach. Within the regression approach, the ANOVA paradigm was considered more appropriate (than the correlational paradigm) because of its greater potential for describing the non-linear use of information.

IMPORTANCE OF THE STUDY

Slovic and Lichtenstein (1971:724) presented a number of generalizations based on an extensive review of research on human judgement:

> First, it is evident that the judge responds in a highly predictable way to the information available to him. Furthermore, much of what is called "intuition" can be explained in a precise and quantitative manner. When this is done, the judge's

> insight into his own cognitive process is found to be inaccurate.
>
> Second, we find that judges have a very difficult time weighting and combining information. . . . To reduce the cognitive strain, they resort to simplified decision strategies, many of which lead them to ignore or misuse relevant information. . . . the structure of the judgement situation is an important determinant of information use.
>
> Finally, despite the great deal of research already completed, it is obvious that we know very little about many aspects of information use in judgement. . . . And the enormous task of integrating this area with the mainstream of cognitive psychology--work on concept formation, problem solving, memory, learning, etc.--remains to be undertaken.

The role of the study was to determine the way in which particular judges responded to five simultaneously presented items of information about proposals for new instructional programs. The judgement of written proposals, of various kinds, has become a major function of educational agencies, at least in Alberta. Examples of such proposals are: proposals for new instructional programs, innovative projects, research projects, and changes in instructional programs.

Knowledge of the way in which judges integrate information, in the evaluation of proposals for new instructional programs, will be of interest to those concerned about that task, particularly the judges themselves. The study will add to the body of research to be integrated with other areas of research, as mentioned in the above quotation. Further, knowledge of how some

educational administrators integrate information, in a particular decision situation, will have implications for the training, of educational administrators, in the area of decision making.

ORGANIZATION OF THE THESIS

The introductory chapter contains: a statement of the purpose of the study; an overview of research on human judgement; a discussion of the importance of the study; and this outline of the structure of the thesis. A review of literature, describing related theory and research, constitutes the second chapter. The third chapter comprises: a statement of the problems investigated; the limitations and delimitations of the study; and a description of the conceptual framework within which the problem was investigated. The fourth chapter contains a discussion of the design of the study. The findings presented in Chapter 5. include: findings related to validity; and the data, and results of the hypothesis testing, for each judge. Chapter 6. contains the conclusions, implications of the findings, and suggestions for further research. The final chapter contains, in summary, descriptions of: the problem, the methodology, the findings, and the conclusions.

Chapter 2

REVIEW OF RELATED LITERATURE

The purpose of this review is to identify variables, which have been studied in relation to human judgement, which were considered in the design of the present study. This chapter includes, for each paradigm: a description, the focal topic, task determinants which have been studied, and conclusions arising from previous research. As discussed previously (page 1.), Slovic and Lichtenstein (1971) was regarded as a major source for this study. Based on an extensive review and comparative analysis of research into human judgement, Slovic and Lichtenstein (1971:653) concluded that much of the recent work has been accomplished within two basic schools of research which they called the "regression" and the "Bayesian" approaches. The present study uses the regression approach. The regression approach was so named because of its characteristic use of multiple regression, and its close relative, analysis of variance, or ANOVA (Slovic and Lichtenstein, 1971:654). Within this broad approach, Slovic and Lichtenstein (1971:655) distinguished two paradigms which they term the

"correlational" paradigm and the "ANOVA" paradigm.

THE CORRELATIONAL PARADIGM

The correlational paradigm has been described as follows (Slovic and Lichtenstein, 1971:644):

> In the correlational paradigm, a judge's integration of information is described by means of correlational statistics. The basic approach requires the judge to make quantitative evaluations of a number of stimuli, each of which is defined by one or more quantified cue dimensions or characteristics. . . . What usually results is that the coefficients of correlation between cues and judgement make public the subtle, and often unreportable, inferential activities of the inferring person. That is, the coefficients reveal the relative degrees that the judgements depend on the various sources of information available to the judge.

The Focal Topic

The focal topic of correlational research has been the modelling of a judge's "policy," the way in which a judge integrates information in making a judgement (Slovic and Lichtenstein, 1971:677-84). Such research has taken the form of: attempting to represent the judge's idiosyncratic weighting policy by means of a linear model (Equation 1.); capturing a judge's objective weighting policy through the use of multiple linear regression analysis; searching for configural usage of cues (in which a judge responds to patterns of two or more cues); and assessing a judge's subjective policy (the relative weights a judge says he

used) and his insight into his own cognitive processes.

The linear model. The linear model, usually used for representing a judge's objective policy, is contained in Equation 1.

$$R = \sum_i b_i X_i \, , \qquad (1)$$

where R is the judge's response, and b(i) and X(i) are, respectively, the weight and value of the *i*th cue.

Slovic and Lichtenstein (1971:677-78) examined more than 30 studies in which researchers attempted to represent the judge's idiosyncratic weighting policy by means of the linear model. The tasks included overall judgements, based on a number of items of information, about: personality characteristics; performance in college; or on the job; attractiveness of common stocks and other types of gambles; physical and mental pathology; and legal matters. In all of these situations, the linear model was said to have done a fairly good job of predicting judgements, as indicated by multiple correlation coefficients in the .80's and .90's for artificial tasks and the .70's for more complex, real-world situations. The use of the ANOVA paradigm, in the present study, provides a further test of the goodness of fit of the linear model and an estimation of the judge's weighting policy through the calculation of the relative magnitudes of

experimental effects.

Capturing a judge's policy. One of the purposes of using a linear model of the judgement process is to make the judge's weighting policy--the way in which the cues determine a judge's responses--explicit (Slovic and Lichtenstein, 1971:678). Within the correlational paradigm, judgement policies are described by means of linear regression equations. The regression weights of the cues defining a stimulus constitute the judge's objective weighting policy. Slovic and Lichtenstein (1971:679) identified five studies which have used regression equations for grouping and clustering judges in terms of the homogeneity of their equations. A description of one such technique, Judgement Analysis, is provided by Houston (1974) who also reports other studies with the same purpose.

Non-linear cue utilization. Despite the high prediction of judgement obtained with the linear model, researchers have continued to search for configural use of cues (where judges respond to patterns of cues). Regression techniques may be made slightly sensitive to the configural use of cues, by including squared ($X1^2$) and cross-product terms ($X1 \times X2$), along with first-order terms ($X1$, $X2$), in the regression equation (Slovic and Lichtenstein, 1971:681).

The ANOVA paradigm, in which interactions are interpreted as evidence for configural use of cues[1] is a statistically efficient mechanism for detecting the configural use of cues. Although a number of studies--in the areas of medical diagnosis, stock market forecasting, and decisions about workmen's compensation cases in a court of law--have shown the existence of configural use of cues, the increment in predictive power contributed by these configural effects was found to be small (Slovic and Lichtenstein, 1971:681). The present study uses the ANOVA paradigm to detect the configural use of cues.

Subjective policies and self-insight. The weights of cues (their relative effect on judgement) determined by the application of regression techniques or ANOVA, are termed objective policies. The relative weights described by the judge, after completing a task, are called his subjective policy. One technique that has been used to determine subjective policies is to ask the judge to distribute 100 points according to the relative importance of each cue[2] (Slovic and Lichtenstein, 1971:683; Slovic, 1969:261). Slovic and Lichtenstein (1971:684) reported

---

[1]A discussion of this interpretation may be found in Chapter 3, Conceptual Framework.

[2]In the present study, each judge was asked to distribute 100 points over the five criteria which were considered most important in the evaluation of program proposals.

that, of the relevant studies reviewed by them, all found serious discrepancies between subjective and objective relative weights. Further, a common error was that judges overestimated the importance they placed on minor cues, and underestimated the importance placed on a few major cues, that is, subjective weights tended to be higher than low objective weights and lower than high objective weights. On the basis of a review of a number of studies, Slovic and Lichtenstein (1971:684) reported that three cues usually sufficed to account for more than 80% of the predictable variance in the judge's responses (overall judgements), the most important cue usually accounting for more than 40% of this variance. The present study used the percentage of the total variance accounted for by experimental effects, as one indicator of the extent to which a judge's cognitive processes could be quantified[1]. Three indicators of a judge's self-insight were used. First, the discrepancy between subjective and objective weights was used as an indicator of a judge's insight into his own cognitive processes. Second, the discrepancy between configural usage of cues identified subjectively and objectively. Third, the discrepancy between the type, interval or ordinal, of stimulus scales identified subjectively and by objective

[1]Described in Chapter 3, Quantification of judgement.

analysis.

Task Determinants and Conclusions

Task determinants[1] which have been studied, using the correlational paradigm, include: (1) cue variability and cue utilization; (2) cue interrelationships; (3) cue format; (4) number of cues; and (5) cue-response compatibility. A discussion of each follows.

Cue variability and cue utilization. Slovic and Lichtenstein (1971:685) reported that two studies--the judgement of IQ on the basis of profiles made up of nine cues; ranking cues on the basis of the combined area of a circle and a square--found that greater weight was placed on a salient cue which had greater variability. The present study used a constant range (high to low) for all cues in all stimuli. Thus, the variability was the same for all cues and was held constant across stimuli.

Cue interrelationships. Slovic and Lichtenstein (1971:685) reported that the intercorrelation and conflict[2] among cues influence the weighting policy used by the judge. The present study used orthogonal (uncorrelated) values in constructing stimuli. Thus, the intercorrelation among cues

---

[1]A task determinant is a characteristic of the experimental task which influences judgement.
[2]Cues were said to conflict when they disagreed in their implications.

was held constant for each judge. Since the five criteria selected by each judge addressed separate issues[1], no conflict among cues was expected.

Cue format. Knox and Hoffman (1962) examined the effect of profile format on judgements of a person's intelligence and sociability. One set of stimuli were presented as T scores with a mean of 50 and standard deviation 10, the second presented the information as percentile scores. The response scale was a 9-point scale with a normal distribution suggested but not forced. The judgements based on the percentile scores were found to be much more variable (a greater range of response scale scores being used) although the regression weights remained stable over the two forms. In the present study, the relative weights of cues were of interest not the magnitudes of responses. A common stimulus scale was used for all stimuli. Thus, differences in cue format did not influence the findings of the present study.

Number of cues. Based on a review of four studies which examined the effect, on judgement, of varying the number of cues presented, Slovic and Lichtenstein (1971:686-87) concluded that:

> . . . there is a small amount of evidence that

---

[1]Discussed in Chapter 4., The Initial Interview.

> increasing the amount of information available to the decision maker increases his confidence without increasing the quality of his decisions and makes his decisions more difficult to predict.

Cue-response compatibility. Based on a review of two studies, which varied the compatibility[1] of cue and response scales, Slovic and Lichtenstein (1971:688) suggested the hypothesis that greater compatibility between a cue and the required response should enhance the importance of the cue in determining the response. In the present study, while different stimulus and response scales were used, the same scales were used in all stimuli. Thus, the hypothesized effect was held constant for all cue-response combinations.

THE ANOVA PARADIGM

A detailed description of the application of the ANOVA paradigm, to the study of human judgement, is a part of the conceptual framework of this study. In return for limiting the factors to relatively few, discrete levels, and making the factors mutually orthogonal, the ANOVA technique provides a statistically efficient mechanism for detecting

---

[1]Compatibility was defined as a function of the similarity between the spatial position of the stimulus in a circular array and the position of a correct response in the same sort of array (Slovic and Lichtenstein, 1971:687). That is, stimulus and response scales are compatible when their metrics are similar.

linear, curvilinear and configural use of information (Slovic and Lichtenstein, 1971:660). Further, within the framework of the ANOVA model, various indices of the importance of individual or joint use of cues, relative to the importance of other cues, can be calculated.

The Focal Topic

The focal topic of the ANOVA paradigm is the search for quantitative models of how sets of cues are integrated into judgements (Slovic and Lichtenstein, 1971:688). The additive model[1] and its variations have received most study. Most of these studies have used rigorous experimental control, factorial designs, and statistical techniques such as ANOVA as the basis of their analyses.

Task Determinants

Task determinants[2], which have been studied within the ANOVA paradigm, include: set size, extremity of information, redundancy, inter-item consistency, and primacy and recency effects. The number of cues in a stimulus has been varied in attempts to distinguish the summation and

---

[1]Represented by Equation 2. in Chapter 3.
[2]A task determinant is a characteristic of the experimental task which influences judgement.

averaging[3] forms of the more general additive model. Varying the combination of extreme and moderate scale scores, within stimuli, provided a means of distinguishing averaging and summative models. Other studies have investigated the relationship between the amount of redundant[1] information and the extremity[2] of the judgement. The degree of consistency among cues has been varied to investigate the effect of inconsistency among cues on the weighting procedure used. Finally, the order of presentation of high and low cue-values, within a stimulus, has been varied to detect the presence of primacy and recency effects. The greater effect of early information is called primacy, and the opposite effect, recency.

<u>Conclusions of Previous Studies</u>

Slovic and Lichtenstein (1971:689-90) report a number of studies of impression formation (where the stimuli were sets of adjectives presented sequentially) designed to distinguish between the summation and averaging models. The

---

[3]In the summation form of the additive model, judges' add the effect of a cue to a judgement based on previously considered cues. The weights of the cues are unrestricted. In the averaging form of the additive model, judges' average the effects of all cues. The weights of cues must sum to one.

[1]A stimulus contains redundant information when cues are related.

[2]A judgement is more extreme when more extreme response scale values are selected.

results of these studies were said to show that sequence of presentation and variation of set size affect judgement. Further, where the results of these studies were claimed to support a summation model, N.H. Anderson (1968) has shown that they can be accounted for by a version of the averaging model. The present study did not have, as a purpose, distinction between summation and averaging models, consequently, the set size was held constant and the five cues constituting a stimulus were presented simultaneously.

A number of studies have varied the extremity of cue dimensions for the purpose of determining the effect of such variation on the weights used in integrating cues. Slovic and Lichtenstein (1971:690) report that the results of these studies have been interpreted as indicative of greater weight being attached to more extreme cues. While estimation of the relative importance of the cues was of concern in the present study, the estimation of the subjective values, assigned by the judge, was not.

Slovic and Lichtenstein (1971:691) reported contradictory findings by studies which varied the amount of redundant information within a stimulus. The degree of redundancy among cues has been measured by: their intercorrelation in a normative sample; and by varying the mixture of cues and combinations of cues in the stimuli (Slovic and Lichtenstein, 1971:691). The stimuli used in the present

study were constructed of five independent criteria with values which were designed to be orthogonal. The same five criteria were used in all stimuli for a particular judge. Consequently, the degree of redundancy was constant for a particular judge.

Slovic and Lichtenstein (1971:691-92) reported some support for the effect, on judgement, of inconsistency among cues. The effects of inconsistency on the weighting of inconsistent cues varies with the nature of the task. In the present study, the context of the experimental task--in which the criteria addressed separate issues, and were therefore independent--was designed to minimize the occurrence of inconsistent cues within stimuli.

Primacy and recency are highly pervasive phenomena, but which effect occurs depends upon the task characteristics (Slovic and Lichtenstein, 1971:693). The present study presented each set of cues simultaneously with the criteria listed in the same order for each stimulus. Whether a judge considered the cues in the order listed or in some other sequence, the effects of primacy and recency were constant for each judge.

SUMMARY

Within the regression approach to the study of human judgement, two paradigms--correlational and ANOVA--have received most attention. The focal topics of the two paradigms are: the goodness of fit of the linear model; capturing a judge's policy; configural use of cues; and subjective policies and self-insight. Task determinants which have been shown to influence judgement, are: cue interrelationships, cue variability; cue format; number of cues; cue-response compatibility; extremity of information; redundancy of information; inter-item consistency; and primacy and recency effects. In the present study, the effects of these variables were controlled, or accounted for, in the experimental design.

Chapter 3

PROBLEM, DELIMITATIONS, AND CONCEPTUAL FRAMEWORK

This chapter contains: a statement of the research problem; the analysis and specification of delimitations; a discussion of the conceptual framework of the study; and the decision criteria for hypothesis testing.

STATEMENT OF THE PROBLEM

The research problem was stated in the fcrm of five hypotheses, to be examined independently, for each judge:

1. The judgement process can be quantified.
2. A judge uses a linear model when integrating five cues which are presented simultaneously.
3. A judge's subjective weighting of criteria will differ from the weighting determined objectively.
4. A judge will not correctly identify the configural use of pairs of cues detected by objective analysis.
5. A judge will not subjectively identify the type of stimulus scale detected by objective analysis.

DELIMITATIONS

The study was delimited in the following ways:

1. the number of cues constituting a stimulus was fixed as five;
2. the cues were presented simultaneously rather than serially;
3. simulated rather than real data, were used in constructing stimuli;
4. actors were encouraged to accept a three-point, ordinal scale for the cues, and a nine-point, interval scale for overall judgement responses;
5. the population was limited to members of Alberta Advanced Education and Manpower who were normally involved in reviewing program proposals, and the sample, to those members who volunteered to participate in the study; and
6. the relationship, of the experimental task to the actual task, was defined for the judges.

The reasons for these delimitations will be presented in Chapter 4. but their implications are relevant here.

Several of the delimitations disallowed generalizations about the real task, evaluation of program proposals, on the basis of findings resulting from the experimental task. Five cues constituted a stimulus, in the experimental

task, whereas the number of cues used in the real task may vary with the judge or the nature of the proposal. Cues were presented simultaneously, in the experimental task, while serial integration of information may be a feature of the real task. The stimuli were not designed to represent those experienced in the actual task of proposal evaluation. Similarly, the context of the experimental task was not defined to represent a real stage in the process of evaluating proposals. The condition that judges consider each stimulus to represent a proposal, and each criterion to have been rated independently of the other four, had no analogue in the real task. Finally, judges were required to respond to stimuli on a nine-point scale, whereas, the response to a real proposal was a decision to fund, or not to fund, a proposed program. Thus, a consequence of the above differences between the real and experimental tasks is that generalizations about the real task cannot be made on the basis of findings about the experimental task.

As discussed previously[1], the study will contribute to the body of knowledge, about human judgement, which has yet to be integrated into the mainstream of cognitive psychology. The small size of the sample, ten, and the non-random selection of actors did not allow generalizations

---

[1]Importance of the Study, Chapter 1.

about human judgement, in general. Since the sample was not a random sample, generalizations to the experimentally accessible population were not attempted.

CONCEPTUAL FRAMEWORK

The first hypothesis, that the judgement process can be quantified, was examined, using a ratio constructed from the estimated magnitudes of experimental effects, the $\underline{w}^2$ coefficient. The second hypothesis, that a judge used a linear model when integrating five cues which are presented simultaneously, was examined within the framework of integration theory. The remaining hypotheses were examined by: comparing that judge's objective and subjective weightings of the five criteria; the configural use of cues identified by the judge with such usage determined by statistical analysis; and the objective and subjective identification of the type of stimulus scales used, interval or ordinal.

Integration Theory

Integration theory, as developed by N.H. Anderson (1968, 1969, 1970), provides a procedure for testing the goodness of fit of linear models of the integration of information by a judge.

The theory. Integration theory has been summarized as follows (Slovic and Lichtenstein, 1971:661):

> Integration theory is concerned simultaneously with two problems. The first is the scaling of stimuli and determining the weighting parameters. This component is called "valuation." The second concern, called "integration," tests theories about the specific composition rules used by subjects.

The second component of the theory, integration, is relevant to the second hypothesis of this study. Research related to integration theory has been concerned with the testing and development of a simple mathematical model for information integration (Anderson, 1968:731).

> It is supposed that the subject has or receives a set of stimuli or pieces of information from which he is to reach some judgement. In very general form, the judgement based on the set is
>
> $$J = C + \sum_{k} w(k)\,s(k), \qquad (2)$$
>
> where J is the judgement, C is a constant, w(k) and s(k) are the value and weight of the kth item in the set (Anderson, 1968:723).

Some assumptions associated with Equation 2. are: that the dependent variable, J, is numeric; and that each stimulus item can be represented by the two parameters s and w, where s represents the value of the item along the dimension of judgement, and w represents the importance of the item in the overall judgement (Anderson, 1968:732). Further, the use of the model requires a theoretically adequate scale of

measurement, the test of goodness of fit provides a joint test of the model and the scale of measurement, and the success of the model establishes the validity of the scale (Anderson, 1968:732).

<u>The test of goodness of fit</u>. With certain assumptions, the test of fit may be reduced to ordinary analysis of variance (ANOVA). The most common of these is the constancy assumption, that the value and weight of any stimulus are essentially the same in all sets (Anderson, 1968:734). On the basis of this constancy assumption, interactions among stimuli will be zero, in theory, and negligible in practice. However, the occurrence of significant interactions does not invalidate the linear model provided that the interaction in concentrated in the multilinear trend component (Anderson, 1967:228). Thus, if the residual of a significant interaction, after partitioning out the bilinear trend component, is not significant, the occurrence of that interaction does not invalidate the linear model. Significant interactions may occur as a result of scale effects rather than as a result of the judge's behaviour.

Elimination of scale effects. The interpretation of interactions is complicated by the possibility of scale effects. Scale effects may be eliminated by experimental precautions, but, as suggested by Anderson (1969:64):

> . . . the work so far largely supports the conclusion that ordinary rating methods, used with moderate experimental care, can provide valid response scales directly. Should invalidity of the response scale be suspected as the cause of failure of a test of fit, a monotone transformation of that scale may save a correct model (Anderson, 1969:64).

Anderson (1970:161) suggests that two possible transformations are of particular interest:

the power function, r=R to the power x, and

the logarithmic function, r=log R,

where r is the scaled value of the response, R. However, transformations of the dependent variable should be based on postulated behaviour laws and the resulting scales would be correct only within the context of the postulated laws (Anderson, 1962:408-10). For example, if a judge reported the use of smaller intervals between higher response scale values, a logarithmic transformation would have been indicated. Conversely, reported usage of smaller intervals between lower response scale values would have supported the use of a power transformation. The transformed responses may be used to validate the model in the same manner as the untransformed responses.

Quantification of Judgement

The magnitudes of experimental effects may be estimated by their omega squared ($\underline{w}^2$) coefficients (Winer, 1971:428-30). The percentage of the total variance accounted for by the experimental effects is, then, the ratio of the sum of the magnitudes of experimental effects to the sum of the magnitudes of all effects underlying an observation. This procedure also allows the estimation of the relative magnitudes, or objective weights, of the main effects. The $\underline{w}^2$ coefficient of a main effect is the ratio of the magnitude of that effect to the sum of the magnitudes of all main effects.

A Judge's Insight

The ANOVA model has the potential for describing linear, curvilinear, and configural aspects of the judgement process as described by Slovic and Lichtenstein (1971:660):

> When judgements are analyzed in terms of an ANOVA model, a significant main effect for cue X1 implies that a judge's responses varied systematically with X1 as the levels of other cues were held constant. If sufficient levels of the factor were included in the design, and if these levels can be assigned interval scale values, the main effects may be divided into effects due to linear, quadratic, cubic, etc. trends. Similarly a significant interaction between cues X1 and X2 implies that the judge was responding to particular patterns of those cues, i.e., the effect of variation of cue X1 upon judgement differed as a function of the corresponding level taken by cue X2.

The use of three factor (criterion) levels allowed the

partitioning of the variance due to a main effect into linear and quadratic components. The variance due to each significant interaction was partitioned into a bilinear trend component and a residual. Given that the evidence supported a linear model of the judgement process (that is, the relative weight of each factor was constant for all levels of the other factors), a significant linear trend component combined with a non-significant quadratic trend component, was interpreted as evidence that the stimulus scale was an interval scale. If the quadratic trend component was significant, then the stimulus scale was only ordinal. In addition, the ANOVA model allows the estimation of the relative importance of experimental effects (as described above). Thus, the ANOVA model provides objective information about four aspects of the judgement process which may be compared to subjective reports, by the judge, on the same aspects. The degree of congruence, between the objective and subjective information, is an indicator of the amount of the judge's insight into his own cognitive processes.

DECISION CRITERIA FOR HYPOTHESIS TESTING

The decision criteria for testing the research hypotheses are discussed below.

Decision Criterion for Hypothesis 1.

Hypothesis 1., that the judgement process can be quantified, was rejected if the percentage of the total variance accounted for by all experimental effects was equal to or less than 80%. This limit was derived from the report of a number of studies which found that 80% of the predictable variance was accounted for by the first three cues (Slovic and Lichtenstein, 1971:684). Since there is some evidence that increasing the number of cues decreases the predictability of the judgement process (Slovic and Lichtenstein, 1971:686-87), the selected limit, 80% of the total variance, reduced the chance of failing to reject the hypothesis when it was false.

Decision Criterion for Hypothesis 2.

Hypothesis 2., that the judge used a linear model for integrating information, was rejected on the occurrence of at least one significant residual, after the bilinear trend component had been partitioned out of the variance due to that interaction. The rationale for this criterion is contained in the conceptual framework of this study. The significance level for testing for the presence of the above effects was set at .05. This level was selected, in preference to the .01 level of significance, to reduce the chance of failing to reject a false null hypothesis. The results of previous studies suggested that the interactions

could be expected to add little to the predictive power of the model contributed by the main effects (Slovic and Lichtenstein, 1971:681). Consequently, the contribution of interactions not significant at the .05 level was expected to be of little practical significance. Thus, the .05 level of significance was preferred to higher levels.

Decision Criterion for Hypothesis 3.

Hypothesis 3., that a judge's subjective weighting of criteria will differ from the weights determined objectively, was rejected if the mean absolute difference, between objective and subjective weights (normalized to 100), exceeded 10 percentage points. This limit was selected arbitrarily to reduce the chance of rejecting a hypothesis when it was true.

Decision Criterion for Hypothesis 4.

Hypothesis 4., that a judge will not correctly identify the configural use of pairs of cues, was not rejected if: a judge failed to subjectively identify all pairs of cues, which were identified, by objective analysis, as used configurally; or, a judge identified more pairs than those identified by objective analysis. Where judges' reported the configural use of three or more cues, the report was examined for possible resolution into the usage of pairs of cues. Where such resolution was not possible, the reported usage was not testable within the design of

this study.

Decision Criterion for Hypothesis 5.

Hypothesis 5. was that a judge will not identify the type of stimulus scale identified by objective analysis. An interval stimulus scale was identified, objectively, by the occurrence, for a criterion, of a statistically significant linear trend component with a quadratic trend component which was not statistically significant. A stimulus scale that was ordinal, but not interval, was identified by the occurrence of a significant quadratic trend component for a criterion. The level of significance, for the above tests, was set at the same level, .05, as for Hypothesis 2., and for the same reasons.

Chapter 4

DESIGN OF THE STUDY

The study uses the conceptual framework described in Chapter 2. to examine five hypotheses about the judgement process of each judge. The hypotheses were:

1. The judgement process can be quantified.
2. A judge uses a linear model when integrating five cues which are presented simultaneously.
3. A judge's subjective weighting of criteria will differ from the weighting determined objectively.
4. A judge will not correctly identify the configural use of pairs of cues detected by objective analysis.
5. A judge will not subjectively identify the type of stimulus scale detected by objective analysis.

This chapter contains operational definitions of some commonly used terms and discusses: the construction of the stimuli; the method of determining a judge's objective and subjective policies; the method of collecting data; and the treatment of the data.

DEFINITIONS

Each _stimulus_ was defined by five cues (a sample stimulus is contained in Appendix 1.).

Each _cue_ consisted of a criterion for evaluating proposals and an associated rating on a three-point scale.

The adjective _subjective_ was used to identify data reported by judges during the interviews.

The adjective _objective_ was used to identify data identified by statistical analysis of the reponses to the experimental task.

CONSTRUCTION OF THE STIMULI

Runkel and McGrath(1972;114-15) identify three elements of a research strategy--precision of measurement, realism of the experimental task, and generalizability of the findings--which cannot be maximized simultaneously. The present study was designed to maximize, primarily, the precision of measurement, and secondarily, the realism of experimental task. The design did not allow the generalization of findings to either the experimentally accessible population or to human judgement, in general.

The context of the judgement task was defined in relation to the actual task of evaluating proposals for new instructional programs. Subjects were required to respond to

each stimulus with an overall judgement on a nine-point scale. Judges were asked to consider the proposals, represented by the stimuli, to have been approved in principle, and thus to have met or exceeded some minimum standard on each criterion. Further, judges were asked to consider each stimulus as representing a program proposal which had been rated, by themselves or others, on each criterion, independently of the other criteria. The purposes of these constraints were: to minimize the conjunctive use of cues; and to establish a link between the simulated stimuli and real proposals.

The Cues

Judges were asked to select appropriate mnemonics for the five distinct criteria which they felt were the most important in the evaluation of program proposals. Each stimulus was then constructed from mnemonics for the five criteria. Each criterion was assigned a rating on a three-point, ordinal scale with values "High," "Medium," and "Low." The combinations of values used in different stimuli were determined through the use of a one-third fractional replication of a 3^5 factorial design, as described by Winer (1971:676-84). Thus, variables considered explicitly in the research design were: the number of cues, the order of presentation of cues, and the combinations of cue values used in constructing stimuli. A discussion of these

considerations follows.

The number of cues. Five was selected, as the number of cues defining a stimulus, for a number of reasons. First, inspection of the departmental policy on program evaluation indicated that about five criteria were all that were readily identifiable. Second, the time required to make one hundred responses was judged, by the author, to be approximate maximum which would be considered reasonable by potential participants in the study. The use of the one-third fractional replicated design required judges to respond to 81 of the 243 possible combinations of factor levels. Third, the number five is within the range recommended by Houston (1974:70) as appropriate for this kind of study.

Fractional replication. The 243 possible combinations of factor levels, for the five criteria defining the stimuli, were divided into three sets using modular arithmetic (Winer, 1971:606-9). The defining relation used was (a+b+c+d+e) mod 3 = i, where a,b,c,d,e are the levels (0,1,or2) of the five factors (the criteria). The set for which i=0 was selected for use in constructing the stimuli of interest. Nine additional combinations were selected randomly from another set (i=2) for use in constructing initial, warm-up, stimuli. The set of 81 combinations was balanced for main effects and 2-way interactions only.

Third- and higher-order interactions were confounded and, hence, pooled to form a residual used as the error term in the analysis of variance. The assumption was made that these third- and higher-order interactions would be negligible. This assumption has been accepted in other studies of human judgement (Slovic and Lichtenstein, 1971:661). The basis of this assumpticn was the lack of supporting evidence, from previous studies, for the configural use of more than two cues. Even where such effects have been found, they have only contributed slightly to the predictive power of the linear model.

Order of presentation of cues. Criteria were assigned, to the factor levels in a stimulus, in the order in which they were identified by a subject. Although the same combinations of factor levels were presented to each judge, the criteria associated with the levels varied from one judge to another. Thus, judges' received different sets of stimuli. The 81 combinations of factor levels were randomly ordered before being used to construct the stimuli, the same order being used for all judges.

The Response

Judges were asked to estimate the strength of the proposals represented by the stimuli and to respond on a nine-point scale with verbal anchors:

weak average strong
1 2 3 4 5 6 7 8 9.

The decision, in the real process of evaluating program proposals, was whether or not to provide funding for the proposed program. Consequently, judges were asked to consider the estimation of the strength of a proposal as a preliminary to the real decision about funding.

The Sample

The experimentally accessible population was defined as professional personnel, in Alberta, who were normally involved in the review of proposals for new instructional programs in higher education. The sample was defined as those members of the Alberta Advanced Education and Manpower who agreed to participate in the study. Ten of the eleven eligible members so agreed.

A Judge's Objective Policy

For the purposes of this study, a judge's *objective policy* was defined as the way in which he integrated information, as discovered through the application of the ANOVA paradigm to the results of the experimental task, in making a judgement. Elements of a judge's objective policy, of interest in this study, were: the extent to which the judgement process could be quantified; the goodness of fit of a linear model; the non-linear use of cues; and the relative weights placed on cues or patterns of cues.

The linear model. The linear additive model was of interest in the present study and no attempt was made to distinguish between the summation and averaging forms of the general model. Under the assumption of constancy of weights across stimuli, the test of goodness of fit of the linear model was reduced to ANOVA. The variation due to each significant interaction was partitioned into its bilinear trend component and a residual (Winer, 1971:478-84). The absence of significant residuals supported the hypothesis that a linear model was used by a judge.

Trend components of main effects. The trend components of significant main effects were computed using the procedure described by Winer (1971:478-84). Main effects were partitioned into their linear and quadratic components. A significant quadratic component was interpreted as indicative of the curvilinear use of a cue, i.e., the use of a stimulus scale which was ordinal, but not interval, for a criterion.

Relative magnitudes of effects. The w^2 coefficient (Winer, 1971:428-30) was used as a measure of the relative magnitudes of experimental effects. This coefficient is a ratio of the magnitude of an effect to the sum of the magnitudes of the effects to which it is being compared. Where the relative magnitude of a combination of effects is required, the numerator of the ratio consists of the sum of

the magnitudes of the combined effects. The omega squared, $\underline{w}^2$, coefficient is interpreted as the proportion of the variance, of the effects included in the denominator, which is accounted for by the effects included in the numerator. Two types of $\underline{w}^2$ coefficient were computed: the proportion of the total variance accounted for by experimental effects, both individually (main effect or interaction) and in groups (main effects, interactions, all effects); and the proportion of the variance, due to the main effects, accounted for by each main effect. The former ratio was used to calculate a measure of the percentage of the total variance accounted for by the experimental effects; and the latter to objectively determine the relative importance of the five criteria in influencing a judge's reponses to the stimuli.

A Judge's Subjective Policy

A judge's _subjective policy_ was defined as the way in which he said that he integrated information in making a judgement. The subjective policy of each judge was represented by three measures; an estimate of the relative importance (weights) of the five criteria in determining the response; the identification, by the judge, of configural usage of cues; and the identification, by the judge, of the stimulus scale as interval or only ordinal.

Subjective weights. The subjective weights of the five criteria were determined by having a judge distribute 100 points over the criteria to represent their importance in determining his responses. This technique was used by Slovic (1969) in a study of the decision making processes of stockbrokers.

Configural use of cues. Judges were asked to identify groups of cues used configurally. *Configural use of cues* was defined, for the judges, as a change in the importance of one or more criteria with changes in the level of another.

Type of stimulus scale. Subjects were provided with a three-point, ordinal stimulus scale with values "High," "Medium," and "Low." Subjects were asked whether the two intervals, between the values subjectively assigned to the three points, were equal or not. Where subjects responded that the two intervals were unequal, an estimate of their relative sizes was requested.

Self-Insight

Hypotheses related to the degree of a judge's insight into his own cognitive processes were examined by the use of three indicators. Where the mean of the absolute differences between the objective and subjective weights, of a particular criterion, exceeded 10 percentage points, a judge's insight into the relative importance of the

criterion, in determining his response, was said to be inaccurate. A judge's insight into his own judgement process was also deemed inaccurate if he failed to subjectively identify those pairs of cues, and only those pairs, identified by ANOVA as configurally used. The subjective identification of configural use of three or more cues could not be tested since third- and higher-order interactions were confounded in the fractional replication design and were pooled to give the error term for the analysis. Subjectively identified usage of the stimulus scale, as interval or not, was compared with such usage identified objectively (a significant quadratic trend component for a main effect was interpreted as indicative of the use of a non-interval scale).

VALIDITY CONSIDERATIONS

Runkel and McGrath (1972) have identified the need to consider the validity of the study design, the data, and the measures.

Validity of the Design

Internal validity of a design is concerned with the relation for which evidence can be presumed, and how repeatable that relation is for the same variables, conditions, and actors (Runkel and McGrath, 1972:169). In the present study, evidence was presumed only for the

relation between cues and responses, that is, the design was internally valid only for that relation. The present study was not cross-validated. Hence, no direct test was made of its repeatability. However, previous studies that have applied the linear model derived from one sample of judgements to predict a second sample have found remarkably little shrinkage in prediction (Slovic and Lichtenstein, 1971:678). This suggests that the linear model would be repeatable for the same variables, conditions, and actors.

External validity of a design is concerned with the breadth and diversity of the population of actors, conditions, and occasions for which the results extracted from the observations can be expected to hold (Runkel and McGrath, 1972:169). The results of the present study were expected to hold for: only those actors studied; and only the conditions used.

Validity of the Data

The internal validity of a datum is concerned with the kinds of statements that can be logically made from the datum (Runkel and McGrath, 1972:285). The assumption was made that a judge's response was based on the cues supplied. The context of the experimental task was designed to maximize the internal validity of the responses. In particular, the use of criteria identified by the judge, and the establishment of a relationship between the real and

experimental tasks, were designed to maximize the internal validity of the data.

The external validity of a datum is concerned with the generality of the datum with respect to: behaviour towards the stimulus; actors; and the dimension of response (Runkel and McGrath, 1972:286-87). The data did not have generality for actors since each actor was presented with a different experimental task, in that different criteria were identified by each judge. The context of the experimental task (in particular, the identification of criteria by the judge, and the establishment of a relationship between the real and experimental tasks) was designed to maximize the generality, for each judge, of behaviour towards the stimuli, and the dimension of response. Further, the nine warm-up stimuli were included to maximize the generality of the data by familiarizing the judge with the experimental task. The presence of significant experimental effects confirmed the external validity of the data for a particular judge. The extent to which the data were externally valid, for a judge, was indicated by: the percentage of the total variance accounted for by the experimental effects; the extent to which a judge considered the experimental task to be "reasonable;" and the occurrence of unlikely combinations of stimulus scale values. Judges were told to consider the experimental task reasonable if they believed that they

could, or did, make meaningful judgements given the proposed stimulus and response scales, and the context of the task.

<u>Validity of the Measures</u>

The validity of a measure is concerned with whether that measure (an operational definition of a property) does indeed measure the conceptually defined property it is intended to measure (Runkel and McGrath, 1972:158). The concept "strength of a proposal" was not defined since its measurement was not of direct interest in the study. The relative importance of experimental effects was defined as a percentage of a variance (total variance or the variance due to main effects). The measure of the relative importance was the omega squared, $\underline{w}^2$, coefficient. The $\underline{w}^2$ coefficients for the effects were consistent (Winer, 1971:429) for each judge. The discrepancy between the objective and subjective weights, for a criterion, was measured by the absolute difference since the direction of the discrepancy was not relevant to Hypothesis 3. The other measures of a judge's subjective policy--subjective reports of configural usage of cues, and the type of stimulus scales used--were assumed to be valid.

DATA COLLECTION

The data were collected in three phases: an initial interview; response to stimuli; and a final interview.

The Initial Interview

The initial interview was used to: explain the purpose of the study; establish the relationship between the real and experimental tasks; determine whether the judge considered the experimental task to be reasonable; identify the mnemonics for five criteria which the judge considered most important in evaluating program proposals; and preview future activities. Each judge was presented with a sample stimulus containing examples of criteria (derived from the departmental Program Coordination Policy) and illustrating the stimulus and response scales. After being informed that response to 90 stimuli would be required, a judge was asked whether he considered the experimental task to be reasonable. Where a judge considered the task reasonable, he was asked to identify the five most important criteria used in the review of program proposals. The criteria were to reflect different elements of a proposal, as outlined in the departmental policy for the review of program proposals. In the event that a judge considered the experimental task unreasonable, he would have been excluded from the sample.

Response to Stimuli

Using the mnemonics for criteria, identified in the initial interview, a set of 90 stimuli were constructed and returned to the judge within three days of the initial interview. The set of stimuli were prefaced by instructions which reiterated the relationship between the real and experimental tasks. A copy of these instructions, together with a sample stimulus, are contained in the Appendix 1. On receipt of a judge's responses, his objective policy was analyzed by the procedures described above.

Final Interview

The final interview was conducted in two parts: collection of data on the judge's subjective policy; and discussion of his objective policy. Each judge was first asked to:

> Distribute 100 points over the five criteria to represent their relative importance in determining the overall evaluation.

Secondly, each judge was asked:

> Do you feel that the rating on any one criterion affected the importance of any other, i.e., did you respond to patterns of two or more cues?

Thirdly, each judge was asked:

> Do you feel that you used equal intervals, High-Medium and Medium-Low, on the stimulus scale? If not, what ratio did you use?

Finally, judges were asked:

> Do you feel, in retrospect, that: the task was

reasonable; and that any particular combination of cues was unrealistic to the extent that it interfered with your judgement?

Results discussed with each judge were: the percentage of the total variance accounted for by all experimental effects; the goodness of fit of the linear model; and the comparison of objective and subjective data on:

1. criterion weights,
2. configural use of cues, and
3. the nature, interval or ordinal, of the stimulus scale.

TREATMENT OF THE DATA

The classical analysis of variance (ANOVA) was carried out using the ANOVA sub-program of the Statistical Package for the Social Sciences, Version 6.02. This sub-program contained the convenient option of pooling third- and higher-order interactions to form the residual which was used as the error term in the analysis. Such pooling was required by the use of fractional replication.

The ANOVA of trend components of experimental effects was performed by a computer program written by the author. The probabilities associated with the F-values were provided by the FISHER subroutine of the XDER library, Division of Educational Research Services, The University of

Alberta. The analyses were performed according to the procedure described by Winer (1971:478-84).

The two forms of $\underline{w}^2$ coefficient were also computed by a computer program written by the author. This program used the procedures described by Winer (1971:428-30). The calculations were based on the F-values of the experimental effects which were obtained from the classical ANOVA.

Chapter 5

FINDINGS

The information required to examine the research hypotheses was: (1) the significance of main effects and interactions detected by classical ANOVA; (2) the significance of the trend components of significant effects; (3) the relative magnitudes of experimental effects; (4) the relative weights of the criteria in determining a judge's responses; (5) the subjective identification of configurally used cues, and (6) the type of stimulus scale used. Since the design of the study did not permit generalizations over judges, the findings are reported separately for each judge. The mnemonics for criteria, selected by each judge, are reported in Appendix 2.

FINDINGS FOR JUDGE ONE

As can be seen from an inspection of the ANOVA summary table (Table 1.), all the main effects were significant (alpha=.05), and none of the interactions were significant (alpha=.05). Since there were no significant interactions, Hypothesis 2.--that Judge One used a linear

Table 1

ANOVA Summary Table for Judge One: Strength of Proposal by Criteria.

Source of Variation	Sum of Square	DF	Mean Square	F	Signif. of F
Main Effects:					
Criterion 1	36.025	2	18.012	17.664	0.001
Criterion 2	8.914	2	4.457	4.371	0.022
Criterion 3	14.247	2	7.123	6.986	0.003
Criterion 4	8.914	2	4.457	4.371	0.022
Criterion 5	91.951	2	45.975	45.085	0.001
Interactions:					
Crit 1 by Crit 2	0.420	4	0.105	0.103	0.981
Crit 1 by Crit 3	0.864	4	0.216	0.212	0.930
Crit 1 by Crit 4	1.309	4	0.327	0.321	0.862
Crit 1 by Crit 5	1.160	4	0.290	0.285	0.886
Crit 2 by Crit 3	3.309	4	0.827	0.811	0.528
Crit 2 by Crit 4	1.086	4	0.272	0.266	0.897
Crit 2 by Crit 5	1.827	4	0.457	0.448	0.773
Crit 3 by Crit 4	2.420	4	0.605	0.593	0.670
Crit 3 by Crit 5	0.494	4	0.123	0.121	0.974
Crit 4 by Crit 5	1.827	4	0.457	0.448	0.773
Explained	174.765	50	3.495		
Residual	30.592	30	1.020		
Total	205.357	80	2.567		

model in integrating the five cues--was not rejected.

An ANOVA summary table for the trend components of significant experimental effects is shown in Table 2. Inspection of this table shows that all the linear trend components were significant (alpha=.05) and all the quadratic trend components were not significant (alpha=.05). Consequently, the stimulus scales were objectively identified as interval scales for all criteria.

The $\underline{w}^2$ coefficients of the experimental effects, for Judge One, are displayed in Table 3. The following percentages of the total variance were accounted for by: main effects, 57.96%; interactions, 10.09%; all effects, 68.05%; residual, 31.95%. Since the percentage of the total variance accounted for by the experimental effects, 68.05%, was less than 80%, Hypothesis 1.--that the judgement process can be quantified--was rejected for Judge One.

The data for comparing, for Judge One, the objective weights, $\underline{w}^2$, and subjective weights of the criteria, are displayed in Table 4. These data include: the relative weights of the criteria; and the differences between their objective and subjective weights. The mean absolute difference, for Judge One, was 19. The largest absolute difference, 30, was found for Criterion 5., and the smallest, 10, for Criterion 4. Since the mean absolute difference was greater than 10 percentage points, Hypothesis

Table 2

ANOVA Summary Table for Judge One: Trend Components for Significant Effects.

Source of Variation	Sum of Square	DF	Mean Square	F	Signif. of F
Criterion 1.					
Linear	34.241	1	34.241	33.569	0.000
Quadratic	1.784	1	1.784	1.749	0.196
Criterion 2.					
Linear	8.167	1	8.167	8.007	0.008
Quadratic	0.747	1	0.747	0.733	0.399
Criterion 3.					
Linear	13.500	1	13.500	13.235	0.001
Quadratic	0.747	1	0.747	0.732	0.399
Criterion 4.					
Linear	8.167	1	8.167	8.007	0.008
Quadratic	0.747	1	0.747	0.733	0.399
Criterion 5.					
Linear	90.741	1	90.741	88.961	0.000
Quadratic	1.210	1	1.210	1.187	0.285
Residual	30.592	30	1.020		

Table 3

Relative Magnitudes of Experimental Effects: Judge One.

Experimental Effect	$\underline{w}^2$
Main Effects:	
Criterion 1.	13.15
Criterion 2.	2.66
Criterion 3.	4.72
Criterion 4.	2.66
Criterion 5.	34.78
Total for Main Effects	57.96
Interactions:	
Crit. 1. by Crit. 2.	1.42
Crit. 1. by Crit. 3.	1.24
Crit. 1. by Crit. 4.	1.07
Crit. 1. by Crit. 5.	1.13
Crit. 2. by Crit. 3.	0.30
Crit. 2. by Crit. 4.	1.16
Crit. 2. by Crit. 5.	0.87
Crit. 3. by Crit. 4.	0.64
Crit. 3. by Crit. 5.	1.39
Crit. 4. by Crit. 5.	0.87
Total for Interactions	10.09
Total for All Effects	68.05
Residual	31.95
Total	100.00

Table 4

Comparison of Objective and Subjective Weights of Criteria: Judge One.

Criterion	Objective Weight	Subjective Weight	Absolute Difference
1	23	5	18
2	5	25	20
3	8	25	17
4	5	15	10
5	60	30	30
Mean Absolute Difference			19

3.--that a judge's subjective weighting of criteria will differ from the weighting determined objectively--was not rejected.

Judge One subjectively identified configural use of Criteria 2., 3., and 5. A value of High, on Criteria 2. and 3., was said to have raised a value of Low on Criterion 5. A value of Low on Criteria 2. and 3., was said to have lowered a value of High on Criterion 5. No pairs of criteria were subjectively identified as being used configurally. Since the configural use of pairs of cues was not identified, either objectively or subjectively, Hypothesis 4.--that a judge will not correctly identify the configural use of pairs of cues detected by objective analysis--was rejected.

Judge One subjectively identified the stimulus scale intervals, High-Medium and Medium-Low, as equal for all cues. The stimulus scales of all cues were identified, objectively, as interval scales. Consequently, Hypothesis 5.--that a judge will not correctly identifiy the type of stimulus scale, identified objectively--was rejected for all criteria.

FINDINGS FOR JUDGE TWO

As can be seen from an inspection of the ANOVA summary table (Table 5.), all the main effects were significant (alpha=.05), and only the interactions between

Table 5

ANOVA Summary Table for Judge Two: Strength of Proposal by Criteria.

Source of Variation	Sum of Square	DF	Mean Square	F	Signif. of F
Main Effects:					
Criterion 1	173.506	2	86.753	119.929	0.001
Criterion 2	9.082	2	4.901	6.776	0.004
Criterion 3	66.691	2	33.346	46.097	0.001
Criterion 4	8.469	2	4.235	5.845	0.007
Criterion 5	7.580	2	3.790	5.240	0.011
Interactions:					
Crit 1 by Crit 2	11.309	4	2.827	3.908	0.011
Crit 1 by Crit 3	24.420	4	6.105	8.440	0.001
Crit 1 by Crit 4	0.642	4	0.160	0.222	0.924
Crit 1 by Crit 5	4.420	4	1.105	1.527	0.219
Crit 2 by Crit 3	1.679	4	0.420	0.580	0.679
Crit 2 by Crit 4	3.457	4	0.864	0.195	0.333
Crit 2 by Crit 5	1.235	4	0.309	0.427	0.788
Crit 3 by Crit 4	0.568	4	0.142	0.196	0.938
Crit 3 by Crit 5	4.568	4	1.142	1.579	0.206
Crit 4 by Crit 5	1.679	4	0.420	0.580	0.679
Explained	320.024	50	6.400		
Residual	21.701	30	0.723		
Total	341.726	80	4.272		

criteria one and two and, one and three, were significant (alpha=.05).

An ANOVA summary table for the trend components of significant experimental effects is displayed in Table 6. For the main effect, due to Criterion 1., both the linear and quadratic trend components were significant (alpha=.05). Consequently, Judge Two used an ordinal, but not interval, stimulus scale for Criterion 1. For the other four criteria, significant linear trend components were found with non-significant quadratic components. Consequently, Judge Two used interval stimulus scales for Criteria 2., 3., 4., and 5. The residual trend components of the two significant interactions--Criterion 1. with Criterion 2., and Criterion 1. with Criterion 3.--were not significant. Consequently, Hypothesis 2.--that a judge uses linear model when integrating five items of simultaneously presented information--was not rejected.

The $\underline{w}^2$ coefficients of the experimental effects, for Judge Two, are displayed in Table 7. The following percentages of the total variance were accounted for by: main effects, 71.94%; interactions, 11.78%; all effects, 83.71%; residual, 16.29%. Since the percentage of the total variance accounted for by the experimental effects, 83.71%, exceeded 80%, Hypothesis 1.--that the judgement process can be quantified--was not rejected.

Table 6

ANOVA Summary Table for Judge Two: Trend Components for Significant Effects.

Source of Variation	Sum of Square	DF	Mean Square	F	Signif. of F
Criterion 1.					
Linear	163.630	1	163.630	226.206	0.000
Quadratic	9.876	1	9.876	13.653	0.001
Criterion 2.					
Linear	9.796	1	9.796	13.543	0.001
Quadratic	0.006	1	0.006	0.008	0.930
Criterion 3.					
Linear	64.463	1	64.463	89.115	0.000
Quadratic	2.228	1	2.228	3.080	0.089
Criterion 4.					
Linear	8.167	1	8.167	11.290	0.002
Quadratic	0.302	1	0.302	0.418	0.523
Criterion 5.					
Linear	5.352	1	5.352	7.399	0.011
Quadratic	2.228	1	2.228	3.080	0.089
Crit 1. by Crit 2.					
Bilinear	5.444	1	5.444	7.527	0.010
Residual	5.865	3	1.955	2.702	0.063
Crit 1. by Crit 3.					
Bilinear	20.250	1	20.250	27.994	0.000
Residual	4.170	3	1.390	1.922	0.147
Residual	21.701	30	0.723		

Table 7

Relative Magnitudes of Experimental Effects: Judge Two.

Experimental Effect	$\underline{w}^2$
Main Effects:	
Criterion 1.	47.82
Criterion 2.	2.32
Criterion 3.	18.13
Criterion 4.	1.95
Criterion 5.	1.70
Total for Main Effects	71.94
Interactions:	
Crit. 1. by Crit. 2.	2.34
Crit. 1. by Crit. 3.	5.98
Crit. 1. by Crit. 4.	0.63
Crit. 1. by Crit. 5.	0.42
Crit. 2. by Crit. 3.	0.34
Crit. 2. by Crit. 4.	0.16
Crit. 2. by Crit. 5.	0.46
Crit. 3. by Crit. 4.	0.65
Crit. 3. by Crit. 5.	0.47
Crit. 4. by Crit. 5.	0.34
Total for Interactions	11.78
Total for All Effects	83.71
Residual	16.29
Total	100.00

The data for comparing for Judge Two, the objective weights, $\underline{w}^2$, and subjective weights of criteria, are displayed in Table 8. These data include: the relative weights of the criteria; and the differences between their objective and subjective weights. The mean absolute difference, for Judge Two, was 14. The largest absolute difference, 31, was found for Criterion 1. The absolute differences for Criteria 3. and 4. were less than 10 percentage points. Since the mean absolute difference, 14, exceeded 10 percentage points, Hypothesis 3.--that a judge's subjective weighting of criteria will differ from the weighting determined objectively--was not rejected.

Judge Two subjectively identified the configural use of criteria as: the value of Criterion 1., Criterion 3., or Criterion 2. (in order of importance) affected the importance of Criterion 4. or Criterion 5. An example was given to illustrate this usage: a High value on Criteria 1., 2., and 3. reduced the importance of Criterion 4. Thus, Judge Two subjectively identified configural usage for two sets of four cues. Since configural use, of pairs of cues was not identified either subjectively or objectively, Hypothesis 4.--that a judge will not correctly identify the configural use of pairs of cues detected by objective analysis--was rejected.

Judge Two subjectively identified the use of

Table 8

Comparison of Objective and Subjective Weights of Criteria: Judge Two.

Criterion	Objective Weight	Subjective Weight	Absolute Difference
1	66	35	31
2	3	20	17
3	25	25	0
4	3	10	7
5	2	15	14
Mean Absolute Difference			16

interval stimulus scales for all criteria. Since a significant quadratic trend component was found for Criterion 1., Hypothesis 5.--that a judge will not subjectively identify the type of stimulus scale detected by objective analysis--was not rejected for that criterion. Since the quadratic trend components of the other four criteria were not significant, Hypothesis 5. was rejected for Criteria 2., 3., 4., and 5.

FINDINGS FOR JUDGE THREE

As can be seen from an inspection of the ANOVA summary table (Table 9.), all the main effects were significant (alpha=.05), and none of the interactions were significant (alpha=.05). As a consequence of the absence of significant interactions, Hypothesis 2.--that a judge uses a linear model when integrating five simultaneously presented cues--was not rejected.

An ANOVA summary table for the trend components of significant experimental effects is shown in Table 10. All the linear trend components of the main effects were significant(alpha=.05), and all their quadratic trend components were not significant(alpha=.05). Thus, Judge Three used an interval stimulus scale for all cues.

The $\underline{w}^2$ coefficients of the experimental effects, for Judge Three, are displayed in Table 11. The following

Table 9

ANOVA Summary Table for Judge Three: Strength of Proposal by Criteria.

Source of Variation	Sum of Square	DF	Mean Square	F	Signif. of F
Main Effects:					
Criterion 1	61.210	2	30.605	70.436	0.001
Criterion 2	22.691	2	11.346	26.112	0.001
Criterion 3	90.839	2	45.420	104.532	0.001
Criterion 4	69.210	2	34.605	79.642	0.001
Criterion 5	21.210	2	10.605	24.407	0.001
Interactions:					
Crit 1 by Crit 2	2.642	4	0.660	1.520	0.221
Crit 1 by Crit 3	2.716	4	0.679	1.563	0.210
Crit 1 by Crit 4	3.235	4	0.809	1.861	0.143
Crit 1 by Crit 5	0.568	4	0.142	0.327	0.858
Crit 2 by Crit 3	1.235	4	0.309	0.710	0.591
Crit 2 by Crit 4	1.086	4	0.272	0.625	0.648
Crit 2 by Crit 5	2.420	4	0.605	1.392	0.260
Crit 3 by Crit 4	1.383	4	0.346	0.796	0.537
Crit 3 by Crit 5	1.160	4	0.290	0.668	0.619
Crit 4 by Crit 5	3.679	4	0.920	2.117	0.103
Explained	285.283	50	5.706		
Residual	13.035	30	0.435		
Total	298.319	80	3.729		

Table 10

ANOVA Summary Table for Judge Three: Trend Components for Significant Effects.

Source of Variation	Sum of Square	DF	Mean Square	F	Signif. of F
Criterion 1.					
Linear	60.167	1	60.167	138.314	0.000
Quadratic	1.043	1	1.043	2.399	0.132
Criterion 2.					
Linear	22.685	1	22.685	52.150	0.000
Quadratic	0.006	1	0.006	0.013	0.909
Criterion 3.					
Linear	90.741	1	90.741	208.599	0.000
Quadratic	0.098	1	0.098	0.226	0.638
Criterion 4.					
Linear	68.907	1	68.907	158.408	0.000
Quadratic	0.303	1	0.303	0.696	0.411
Criterion 5.					
Linear	20.167	1	20.167	46.360	0.000
Quadratic	1.043	1	1.043	2.399	0.132
Residual	13.035	30	0.435		

Table 11

Relative Magnitudes of Experimental Effects: Judge Three.

Experimental Effect	$\underline{w}^2$
Main Effects:	
Criterion 1.	19.77
Criterion 2.	7.15
Criterion 3.	29.47
Criterion 4.	22.39
Criterion 5.	6.66
Total for Main Effects	85.44
Interactions:	
Crit. 1. by Crit. 2.	0.30
Crit. 1. by Crit. 3.	0.32
Crit. 1. by Crit. 4.	0.49
Crit. 1. by Crit. 5.	0.38
Crit. 2. by Crit. 3.	0.17
Crit. 2. by Crit. 4.	0.21
Crit. 2. by Crit. 5.	0.22
Crit. 3. by Crit. 4.	0.12
Crit. 3. by Crit. 5.	0.19
Crit. 4. by Crit. 5.	0.64
Total for Interactions	3.03
Total for All Effects	88.47
Residual	11.53
Total	100.00

percentages of the total variance were accounted for by: main effects, 85.44%; interactions, 3.03%; all effects, 88.47%; residual, 11.53%. Since the percentage of the total variance accounted for by all experimental effects, 88.47%, exceeded 80%, Hypothesis 1.--that the judgement process can be quantified--was not rejected.

The data for comparing, for Judge Three, the objective weights, $\underline{w}^2$, and subjective weights of the criteria, are displayed in Table 12. The mean absolute difference was 4 percentage points. The largest absolute difference, 10, was found for Criterion 3., and the smallest, 1, for Criterion 4. Thus, all the absolute differences did not exceed 10 percentage points. Since the mean absolute difference, 4, did not exceed 10 percentage points, Hypothesis 3.--that a judge's subjective weighting of criteria will differ from the weighting identified objectively--was rejected.

Since Judge Three subjectively identified no pairs of cues as being used configurally, and none were detected by objective analysis, Hypothesis 4.--that a judge will not correctly identify the configural use of pairs of cues--was rejected.

Judge Three subjectively identified the stimulus scales for all criteria as interval scales. Since objective analysis identified the stimulus scales for all criteria as

Table 12

Comparison of Objective and Subjective Weights of Criteria: Judge Three.

Criterion	Objective Weight	Subjective Weight	Absolute Difference
1	23	25	2
2	8	15	7
3	35	25	10
4	26	25	1
5	8	10	2
Mean Absolute Difference			4

interval scales, Hypothesis 5.--that a judge will not subjectively indentify type of stimulus scale identified objectively--was rejected for all criteria.

FINDINGS FOR JUDGE FOUR

As can be seen from an inspection of the ANOVA summary table (Table 13.), all the main effects were significant (alpha=.05), and none of the interactions were significant (alpha=.05). As a consequence of the absence of a significant interactions, Hypothesis 2.--that a judge uses a linear model when integrating five simultaneously presented cues--was not rejected.

An ANOVA summary table for the trend components of significant experimental effects is shown in Table 14. The linear trend components were significant (alpha=.05) for all main effects. The only significant (alpha=.05) quadratic trend component was found for Criterion 5. Thus, only the stimulus scale for Criterion 5. was an ordinal, but not interval, scale, the scales for the other four criteria being interval scales.

The $\underline{w}^2$ coefficients of the experimental effects for Judge Four are displayed in Table 15. The following percentages of the total variance were accounted for by: main effects, 84.45%; interactions, 2.16%; all effects, 87.61%; residual, 12.39%. Since the percentage of the total

Table 13

ANOVA Summary Table for Judge Four: Strength of Proposal by Criteria.

Source of Variation	Sum of Square	DF	Mean Square	F	Signif. of F
Main Effects:					
Criterion 1	98.988	2	49.494	120.038	0.001
Criterion 2	20.173	2	10.086	24.463	0.001
Criterion 3	62.914	2	31.457	76.293	0.001
Criterion 4	39.210	2	19.605	47.548	0.001
Criterion 5	13.136	2	6.568	15.929	0.001
Interactions:					
Crit 1 by Crit 2	2.346	4	0.586	1.422	0.251
Crit 1 by Crit 3	1.383	4	0.346	0.838	0.512
Crit 1 by Crit 4	0.864	4	0.216	0.524	0.719
Crit 1 by Crit 5	1.160	4	0.290	0.704	0.596
Crit 2 by Crit 3	0.864	4	0.216	0.524	0.719
Crit 2 by Crit 4	0.790	4	0.198	0.479	0.751
Crit 2 by Crit 5	1.531	4	0.383	0.928	0.461
Crit 3 by Crit 4	1.827	4	0.457	1.108	0.371
Crit 3 by Crit 5	2.123	4	0.531	1.288	0.297
Crit 4 by Crit 5	0.494	4	0.123	0.299	0.876
Explained	247.802	50	4.956		
Residual	12.370	30	0.412		
Total	260.172	80	3.252		

Table 14

ANOVA Summary Table for Judge Four: Trend Components for Significant Effects.

Source of Variation	Sum of Square	DF	Mean Square	F	Signif. of F
Criterion 1.					
Linear	98.685	1	98.685	239.334	0.000
Quadratic	0.303	1	0.303	0.734	0.398
Criterion 2.					
Linear	18.963	1	18.963	45.989	0.000
Quadratic	1.210	1	1.210	2.935	0.097
Criterion 3.					
Linear	62.296	1	62.296	151.082	0.000
Quadratic	0.618	1	0.618	1.498	0.230
Criterion 4.					
Linear	39.185	1	39.185	95.033	0.000
Quadratic	0.025	1	0.025	0.060	0.808
Criterion 5.					
Linear	10.667	1	10.667	25.869	0.000
Quadratic	2.469	1	2.469	5.989	0.020
Residual	12.370	30	0.412		

Table 15

Relative Magnitudes of Experimental Effects: Judge Four.

Experimental Effect	$\underline{w}^2$
Main Effects:	
Criterion 1.	36.42
Criterion 2.	7.18
Criterion 3.	23.04
Criterion 4.	14.24
Criterion 5.	4.57
Total for Main Effects	85.45
Interactions:	
Crit. 1. by Crit. 2.	0.26
Crit. 1. by Crit. 3.	0.10
Crit. 1. by Crit. 4.	0.29
Crit. 1. by Crit. 5.	0.18
Crit. 2. by Crit. 3.	0.29
Crit. 2. by Crit. 4.	0.32
Crit. 2. by Crit. 5.	0.04
Crit. 3. by Crit. 4.	0.07
Crit. 3. by Crit. 5.	0.18
Crit. 4. by Crit. 5.	0.43
Total for Interactions	2.16
Total for All Effects	87.61
Residual	12.39
Total	100.00

variance accounted for by all experimental effects, 87.61%, exceeded 80%, Hypothesis 1.--that the judgement process can be quantified--was not rejected.

The data for comparing, for Judge Four, the objective weights, $\underline{w}^2$, and subjective weights of the criteria, are displayed in Table 16. The mean absolute difference was 5 percentage points. The largest absolute difference, 8, was found for Criterion 1., and the smallest, 2, for Criteria 3. and 4. Since the mean absolute difference, 5, did not exceed 10 percentage points, Hypothesis 3.--that a judge's subjective weighting of criteria will differ from the weighting identified objectively--was rejected.

Judge Four reported that the values of Criteria 1. and 3. affected the importance of the other criteria: if Criteria 1. and 3. were High, the other criteria were less important than for other values of Criteria 1. and 3. Further, Cues 2., 4., and 5. were used to adjust an initial estimate based on Cues 1. and 3. Thus, Judge Four did not identify the configural use of pairs of cues. Since no configural usage of pairs of cues was identified by objective analysis, Hypothesis 4.--that a judge will not correctly identify the configural usage of pairs of cues--was rejected.

Judge Four subjectively identified the use of

Table 16

Comparison of Objective and Subjective Weights of Criteria: Judge Four.

Criterion	Objective Weight	Subjective Weight	Absolute Difference
1	43	35	8
2	8	15	7
3	27	25	2
4	17	15	2
5	5	10	5
Mean Absolute Difference			5

interval stimulus scales for all criteria. However, objective analysis showed the use of an ordinal stimulus scale for Criterion 5. and interval stimulus scales for the remaining criteria. Thus, Hypothesis 5.--that a judge will not subjectively identify the type of stimulus scale identified objectively--was rejected for Criteria 1. through 4., but was not rejected for Criterion 5.

FINDINGS FOR JUDGE FIVE

As can be seen from an inspection of the ANOVA summary table (Table 17.), all the main effects were significant (alpha=.05), and none of the interactions were significant (alpha=.05). As a consequence of the absence of significant interactions, Hypothesis 2.--that a judge uses a linear model when integrating five simultaneously presented cues--was not rejected.

An ANOVA summary table for the trend components of significant experimental effects is shown in Table 18. All the linear and quadratic trend components of the main effects were found to be significant (alpha=.05). Thus, Judge Five used an ordinal, but not interval, stimulus scale for all criteria.

The $\underline{w}^2$ coefficients of the experimental effects, for Judge Five, are displayed in Table 19. The following percentages of the total variance were accounted for by:

Table 17

ANOVA Summary Table for Judge Five: Strength of Proposal by Criteria.

Source of Variation	Sum of Square	DF	Mean Square	F	Signif. of F
Main Effects:					
Criterion 1	27.951	2	13.975	63.600	0.001
Criterion 2	55.728	2	27.864	126.807	0.001
Criterion 3	46.099	2	23.049	104.896	0.001
Criterion 4	28.173	2	14.086	64.106	0.001
Criterion 5	49.654	2	24.827	112.986	0.001
Interactions:					
Crit 1 by Crit 2	0.790	4	0.198	0.899	0.477
Crit 1 by Crit 3	0.420	4	0.105	0.478	0.752
Crit 1 by Crit 4	1.457	4	0.364	1.657	0.186
Crit 1 by Crit 5	0.864	4	0.216	0.983	0.432
Crit 2 by Crit 3	2.198	4	0.549	2.500	0.063
Crit 2 by Crit 4	0.790	4	0.198	0.899	0.477
Crit 2 by Crit 5	0.198	4	0.049	0.225	0.922
Crit 3 by Crit 4	0.198	4	0.049	0.225	0.922
Crit 3 by Crit 5	0.272	4	0.068	0.309	0.870
Crit 4 by Crit 5	0.864	4	0.216	0.983	0.432
Explained	215.654	50	4.313		
Residual	6.592	30	0.220		
Total	222.246	80	2.778		

Table 18

ANOVA Summary Table for Judge Five: Trend Components for Significant Effects.

Source of Variation	Sum of Square	DF	Mean Square	F	Signif. of F
Criterion 1.					
Linear	26.741	1	26.741	121.696	0.000
Quadratic	1.210	1	1.210	5.508	0.026
Criterion 2.					
Linear	48.167	1	48.167	219.205	0.000
Quadratic	7.561	1	7.561	34.411	0.000
Criterion 3.					
Linear	40.907	1	40.907	186.168	0.000
Quadratic	5.192	1	5.192	23.627	0.000
Criterion 4.					
Linear	24.000	1	24.000	109.223	0.000
Quadratic	4.173	1	4.173	18.991	0.000
Criterion 5.					
Linear	44.463	1	44.463	202.350	0.000
Quadratic	5.191	1	5.191	23.624	0.000
Residual	6.592	30	0.220		

Table 19

Relative Magnitudes of Experimental Effects: Judge Five.

Experimental Effect	w^2
Main Effects:	
Criterion 1.	12.08
Criterion 2.	24.28
Criterion 3.	20.05
Criterion 4.	12.18
Criterion 5.	21.61
Total for Main Effects	90.19
Interactions:	
Crit. 1. by Crit. 2.	0.04
Crit. 1. by Crit. 3.	0.20
Crit. 1. by Crit. 4.	0.25
Crit. 1. by Crit. 5.	0.01
Crit. 2. by Crit. 3.	0.58
Crit. 2. by Crit. 4.	0.04
Crit. 2. by Crit. 5.	0.30
Crit. 3. by Crit. 4.	0.30
Crit. 3. by Crit. 5.	0.27
Crit. 4. by Crit. 5.	0.01
Total for Interactions	1.99
Total for All Effects	92.18
Residual	7.82
Total	100.00

main effects, 90.19%; interactions, 1.99%; all effects, 92.18%; residual, 7.82%. Since the percentage of the total variance accounted for by all experimental effects, 92.18%, exceeded 80%, Hypothesis 1.--that the judgement process can be quantified--was not rejected.

The data for comparing, for Judge Five, the objective weights, $\underline{w}^2$, and subjective weights of the criteria, are displayed in Table 20. The mean absolute difference was 2 percentage points. The largest absolute difference, 4, was found for Criterion 5., and the smallest, 0, for Criterion 2. Thus, all the absolute differences were less than 10 percentage points. Since the mean absolute difference, 2, did not exceed 10 percentage points, Hypothesis 3.--that a judge's subjective weighting of criteria will differ from the weighting identified objectively--was rejected.

Since Judge Five subjectively identified no pairs of cues as being used configurally, and none were detected by objective analysis, Hypothesis 4.--that a judge will not correctly identify the configural use of pairs of cues--was rejected.

Judge Five subjectively identified the stimulus scales for all criteria as interval scales. Since objective analysis identified the stimulus scales for all criteria as ordinal, but not interval scales, Hypothesis 5.--that a

Table 20

Comparison of Objective and Subjective Weights of Criteria: Judge Five.

Criterion	Objective Weight	Subjective Weight	Absolute Difference
1	13	15	2
2	27	27	0
3	22	23	1
4	14	15	1
5	24	20	4
Mean Absolute Difference			2

judge will not subjectively indentify type of stimulus scale identified objectively--was not rejected for any criterion.

FINDINGS FOR JUDGE SIX

As can be seen from an inspection of the ANOVA summary table (Table 21.), all the main effects were significant (alpha=.05), and none of the interactions were significant (alpha=.05). As a consequence of the absence of a significant interactions, Hypothesis 2.--that a judge uses a linear model when integrating five simultaneously presented cues--was not rejected.

An ANOVA summary table for the trend components of significant experimental effects is shown in Table 22. The linear trend components were significant (alpha=.05) for all main effects. The only significant (alpha=.05) quadratic trend component was found for Criterion 5. Thus, the stimulus scale for Criterion 5. was ordinal, but not interval, the scales for the other four criteria being interval.

The $\underline{w}^2$ coefficients of the experimental effects for Judge Six are displayed in Table 23. The following percentages of the total variance were accounted for by: main effects, 75.60%; interactions, 5.05%; all effects, 80.65%; residual, 19.35%. Since the percentage of the total variance accounted for by all experimental effects, 80.65%,

Table 21

ANOVA Summary Table for Judge Six: Strength of Proposal by Criteria.

Source of Variation	Sum of Square	DF	Mean Square	F	Signif. of F
Main Effects:					
Criterion 1	44.469	2	22.235	55.933	0.001
Criterion 2	20.173	2	10.086	25.373	0.001
Criterion 3	6.617	2	3.309	8.323	0.001
Criterion 4	17.802	2	8.901	22.392	0.001
Criterion 5	40.691	2	20.346	51.182	0.001
Interactions:					
Crit 1 by Crit 2	0.790	4	0.198	0.497	0.738
Crit 1 by Crit 3	1.235	4	0.309	0.776	0.549
Crit 1 by Crit 4	0.938	4	0.235	0.590	0.672
Crit 1 by Crit 5	4.049	4	1.012	2.547	0.060
Crit 2 by Crit 3	2.420	4	0.605	1.522	0.221
Crit 2 by Crit 4	1.679	4	0.420	1.056	0.395
Crit 2 by Crit 5	0.568	4	0.142	0.357	0.837
Crit 3 by Crit 4	2.790	4	0.698	1.755	0.164
Crit 3 by Crit 5	1.901	4	0.475	1.196	0.333
Crit 4 by Crit 5	2.272	4	0.568	1.429	0.249
Explained	148.395	50	2.968		
Residual	11.926	30	0.398		
Total	160.320	80	2.004		

Table 22

ANOVA Summary Table for Judge Six: Trend Components for Significant Effects.

Source of Variation	Sum of Square	DF	Mean Square	F	Signif. of F
Criterion 1.					
Linear	44.463	1	44.463	111.716	0.000
Quadratic	0.033	1	0.033	0.083	0.775
Criterion 2.					
Linear	20.167	1	20.167	50.670	0.000
Quadratic	0.006	1	0.006	0.016	0.900
Criterion 3.					
Linear	6.000	1	6.000	15.075	0.001
Quadratic	0.617	1	0.617	1.550	0.223
Criterion 4.					
Linear	17.796	1	17.796	44.714	0.000
Quadratic	0.006	1	0.006	0.014	0.905
Criterion 5.					
Linear	35.852	1	38.852	90.080	0.000
Quadratic	4.839	1	4.839	12.159	0.002
Residual	11.926	30	0.398		

Table 23

Relative Magnitudes of Experimental Effects: Judge Six.

Experimental Effect	$\underline{w}^2$
Main Effects:	
Criterion 1.	26.25
Criterion 2.	11.65
Criterion 3.	3.50
Criterion 4.	10.22
Criterion 5.	23.98
Total for Main Effects	75.60
Interactions:	
Crit. 1. by Crit. 2.	0.48
Crit. 1. by Crit. 3.	0.21
Crit. 1. by Crit. 4.	0.39
Crit. 1. by Crit. 5.	1.48
Crit. 2. by Crit. 3.	0.50
Crit. 2. by Crit. 4.	0.05
Crit. 2. by Crit. 5.	0.61
Crit. 3. by Crit. 4.	0.72
Crit. 3. by Crit. 5.	0.19
Crit. 4. by Crit. 5.	0.41
Total for Interactions	5.05
Total for All Effects	80.65
Residual	19.35
Total	100.00

exceeded 80%, Hypothesis 1.--that the judgement process can be quantified--was not rejected.

The data for comparing, for Judge Six, the objective weights, $\underline{w}^2$, and subjective weights of criteria, are displayed in Table 24. These data include the relative weights of the criteria and the absolute differences between their objective and subjective weights. The mean absolute difference was 9 percentage points. The largest absolute difference, 12, was found for Criterion 5., and the smallest, 5, for Criterion 1. The absolute differences for four of the five criteria did not exceed 10 percentage points. Since the mean absolute difference, 9, did not exceed 10 percentage points, Hypothesis 3.--that a judge's subjective weighting of criteria will differ from the weighting determined objectively--was rejected.

Judge Six subjectively identified the following combinations of criteria as being used configurally: Criterion 1. with Criterion 3. or Criterion 5.; and Criterion 2. with Criterion 3. and Criterion 5. Thus, Judge Six subjectively identified configural usage for two pairs of criteria: Criteria 1. and 3. or 1. and 5. Since no configural usage of pairs of cues was detected by objective analysis, Hypothesis 4.--that a judge will not correctly identify the configural use of pairs of cues detected by objective analysis--was not rejected.

Table 24

Comparison of Objective and Subjective Weights of Criteria: Judge Six.

Criterion	Objective Weight	Subjective Weight	Absolute Difference
1	35	30	5
2	15	25	10
3	5	15	10
4	14	20	6
5	32	10	12
Mean Absolute Difference			9

Judge Six subjectively identified the stimulus scales used for all criteria as ordinal, but not interval scales. The interval Medium-Low was said to be twice the interval High-Medium. Since the stimulus scales for Criteria 1., 2., 3., and 4. were identified, by objective analysis, as interval scales, Hypothesis 5.--that a judge will not subjectively identify the type of stimulus scale detected by objective analysis--was not rejected for these four criteria. Since the stimulus scale for Criterion 5. was identified objectively as ordinal, but not interval, Hypothesis 5. was rejected for this criterion.

FINDINGS FOR JUDGE SEVEN

As can be seen from an inspection of the ANOVA summary table (Table 25.), all main effects were significant (alpha=.05), except for Criterion 3. Three significant (alpha=.05) interactions were found: Criterion 1. with Criterion 2., Criterion 1. with Criterion 4., and Criterion 4. with Criterion 5.

An ANOVA summary table for the trend components of significant experimental effects is shown in Table 26. The linear trend components of all the main effects were significant (alpha=.05) while their quadratic trend components were not. Thus, interval stimulus scales were used for all criteria. The bilinear trend components of all

Table 25

ANOVA Summary Table for Judge Seven: Strength of Proposal by Criteria.

Source of Variation	Sum of Square	DF	Mean Square	F	Signif. of F
Main Effects:					
Criterion 1	137.062	2	68.531	81.404	0.001
Criterion 2	65.210	2	32.605	38.730	0.001
Criterion 3	1.580	2	0.790	0.939	0.402
Criterion 4	73.654	2	36.827	43.745	0.001
Criterion 5	9.951	2	4.975	5.910	0.007
Interactions:					
Crit 1 by Crit 2	22.938	4	5.735	6.812	0.001
Crit 1 by Crit 3	6.568	4	1.642	1.950	0.128
Crit 1 by Crit 4	19.161	4	4.790	5.690	0.002
Crit 1 by Crit 5	4.642	4	1.160	1.378	0.265
Crit 2 by Crit 3	6.642	4	1.660	1.972	0.124
Crit 2 by Crit 4	1.901	4	0.475	0.565	0.690
Crit 2 by Crit 5	1.383	4	0.346	0.411	0.800
Crit 3 by Crit 4	4.197	4	1.049	1.246	0.313
Crit 3 by Crit 5	1.235	4	0.309	0.367	0.830
Crit 4 by Crit 5	14.272	4	3.568	4.238	0.008
Explained	370.395	50	7.408		
Residual	25.256	30	0.842		
Total	395.651	80	4.946		

Table 26

ANOVA Summary Table for Judge Seven: Trend Components for Significant Effects.

Source of Variation	Sum of Square	DF	Mean Square	F	Signif. of F
Criterion 1.					
Linear	133.796	1	133.796	158.928	0.000
Quadratic	3.266	1	3.266	3.879	0.058
Criterion 2.					
Linear	64.463	1	64.463	76.571	0.000
Quadratic	0.747	1	0.747	0.887	0.354
Criterion 4.					
Linear	73.500	1	73.500	87.306	0.000
Quadratic	0.154	1	0.154	0.183	0.672
Criterion 5.					
Linear	8.167	1	8.167	9.701	0.004
Quadratic	1.784	1	1.784	2.119	0.156
Crit 1. by Crit 2.					
Bilinear	20.250	1	20.250	24.054	0.000
Residual	2.688	3	0.896	1.064	0.379
Crit 1. by Crit 4.					
Bilinear	14.694	1	14.694	17.455	0.000
Residual	4.467	3	1.489	1.769	0.174
Crit 4. by Crit 5.					
Bilinear	7.111	1	7.111	8.447	0.007
Residual	7.161	3	2.387	2.835	0.055
Residual	25.256	30	0.842		

significant interactions were all significant while their residuals were not. As a consequence of the absence of significant residuals of significant interactions, Hypothesis 2.--that a judge uses a linear model when integrating five simultaneously presented cues--was not rejected.

The $\underline{w}^2$ coefficients of the experimental effects for Judge Seven are displayed in Table 27. The following percentages of the total variance were accounted for by: main effects, 68.42%, interactions, 14.84%; all effects, 83.26%; residual, 16.74%. Since the percentage of the total variance accounted for by all the experimental effects, 83.26%, exceeded 80%, Hypothesis 1.--that the judgement process can be quantified--was not rejected.

The data for comparing, for Judge Seven, the objective weights, $\underline{w}^2$, and the subjective weights of criteria, are displayed in Table 28. These data include the relative weights of criteria and the absolute differences between their objective and subjective weights. The mean absolute difference was 10 percentage points. The largest absolute difference, 14, was found for Criterion 1., and the smallest, 2, for Criterion 2. Only two of the absolute differences were less than or equal to 10 percentage points. Since the mean absolute difference, 10, did not exceed 10 percentage points, Hypothesis 3.--that a judge's subjective

Table 27

Relative Magnitudes of Experimental Effects: Judge Seven.

Experimental Effect	$\underline{w}^2$
Main Effects:	
Criterion 1.	33.24
Criterion 2.	15.45
Criterion 3.	0.03
Criterion 4.	17.67
Criterion 5.	2.03
Total for Main Effects	68.42
Interactions:	
Crit. 1. by Crit. 2.	4.81
Crit. 1. by Crit. 3.	0.79
Crit. 1. by Crit. 4.	3.88
Crit. 1. by Crit. 5.	0.31
Crit. 2. by Crit. 3.	0.80
Crit. 2. by Crit. 4.	0.36
Crit. 2. by Crit. 5.	0.49
Crit. 3. by Crit. 4.	0.20
Crit. 3. by Crit. 5.	0.52
Crit. 4. by Crit. 5.	2.68
Total for Interactions	14.84
Total for All Effects	83.26
Residual	16.74
Total	100.00

Table 28

Comparison of Objective and Subjective Weights of Criteria: Judge Seven.

Criterion	Objective Weight	Subjective Weight	Absolute Difference
1	49	35	14
2	23	25	2
3	0	10	10
4	26	15	11
5	3	15	12
Mean Absolute Difference			10

weighting of criteria will differ from the weighting determined objectively--was rejected.

Judge Seven identified the following pairs of criteria as being used configurally: Criterion 1. with Criterion 2.; Criterion 2. with Criterion 4.; and Criterion 1. with Criterion 4. (listed in order of importance). Since objective analysis detected no configural usage of pairs of cues, Hypothesis 4.--that a judge will not correctly identify the configural use of pairs of cues detected by objective analysis--was not rejected.

Judge Seven reported that the type of stimulus scale used, for a criterion, varied with the stimulus and depended on the impression gained from other cues. Since objective analysis detected the use of an interval stimulus scale, for all criteria, Hypothesis 5.--that a judge will not subjectively identify the type of stimulus scale detected by objective analysis--was not rejected for all criteria.

FINDINGS FOR JUDGE EIGHT

As can be seen from an inspection of the ANOVA summary table (Table 29.), all the main effects were significant (alpha=.05) and none of the interactions were significant (alpha=.05). As a consequence of the absence of significant interactions, Hypothesis 2.--that a judge uses a linear model when integrating five simultaneously presented

Table 29

ANOVA Summary Table for Judge Eight: Strength of Proposal by Criteria.

Source of Variation	Sum of Square	DF	Mean Square	F	Signif. of F
Main Effects:					
Criterion 1	26.469	2	13.235	67.007	0.001
Criterion 2	36.321	2	18.160	91.947	0.001
Criterion 3	21.432	2	10.716	54.256	0.001
Criterion 4	11.580	2	5.790	29.316	0.001
Criterion 5	17.358	2	8.679	43.942	0.001
Interactions:					
Crit 1 by Crit 2	1.605	4	0.401	2.031	0.115
Crit 1 by Crit 3	1.161	4	0.290	1.469	0.236
Crit 1 by Crit 4	0.346	4	0.086	0.438	0.780
Crit 1 by Crit 5	1.012	4	0.253	1.281	0.299
Crit 2 by Crit 3	1.975	4	0.494	2.500	0.063
Crit 2 by Crit 4	0.716	4	0.179	0.906	0.473
Crit 2 by Crit 5	1.605	4	0.401	2.031	0.115
Crit 3 by Crit 4	0.049	4	0.012	0.063	0.992
Crit 3 by Crit 5	0.494	4	0.123	0.625	0.648
Crit 4 by Crit 5	1.679	4	0.420	2.125	0.102
Explained	123.802	50	2.476		
Residual	5.925	30	0.198		
Total	129.728	80	1.622		

items of information--was not rejected.

An ANOVA summary table for the trend components of significant experimental effects is shown in Table 30. Both the linear and quadratic trend components were significant (alpha=.05) for all main effects. Thus, Judge Eight used an ordinal, but not interval, scale for all criteria.

The $\underline{w}^2$ coefficients of the experimental effects for Judge Eight are displayed in Table 31. The following percentages of the total variance were accounted for by: main effects, 83.58%; interactions, 4.40%; all effects, 87.97%; residual, 12.03%. Since the percentage of the total variance accounted for by all the experimental effects, 83.58%, exceeded 80%, Hypothesis 1.--that the judgement process can be quantified--was not rejected.

Data for comparing, for Judge Eight, the objective weights, $\underline{w}^2$, and subjective weights of criteria, are displayed in Table 32. These data include the relative weights of criteria and the absolute differences between their objective and subjective weights. The mean absolute difference was 2 percentage points. The largest absolute difference, 5, was found for Criterion 5., and the smallest, 0, for Criterion 4. Since the mean absolute difference, 2, did not exceed 10 percentage points, Hypothesis 3.--that a judge's subjective weighting of criteria will differ from the weighting determined objectively--was rejected.

Table 30

ANOVA Summary Table for Judge Eight: Trend Components for Significant Effects.

Source of Variation	Sum of Square	DF	Mean Square	F	Signif. of F
Criterion 1.					
Linear	24.000	1	24.000	121.212	0.000
Quadratic	2.469	1	2.469	12.470	0.001
Criterion 2.					
Linear	31.130	1	31.130	157.220	0.000
Quadratic	5.101	1	5.101	25.765	0.000
Criterion 3.					
Linear	18.963	1	18.963	95.772	0.000
Quadratic	2.469	1	2.469	12.470	0.001
Criterion 4.					
Linear	9.796	1	9.796	49.476	0.000
Quadratic	1.784	1	1.784	9.009	0.005
Criterion 5.					
Linear	15.574	1	15.574	78.657	0.000
Quadratic	1.784	3	1.784	9.010	0.005
Residual	5.925	30	0.198		

Table 31

Relative Magnitudes of Experimental Effects: Judge Eight.

Experimental Effect	$\underline{w}^2$
Main Effects:	
Criterion 1.	19.60
Criterion 2.	27.01
Criterion 3.	15.81
Criterion 4.	8.41
Criterion 5.	12.75
Total for Main Effects	83.58
Interactions:	
Crit. 1. by Crit. 2.	0.61
Crit. 1. by Crit. 3.	0.28
Crit. 1. by Crit. 4.	0.33
Crit. 1. by Crit. 5.	0.17
Crit. 2. by Crit. 3.	0.89
Crit. 2. by Crit. 4.	0.06
Crit. 2. by Crit. 5.	0.61
Crit. 3. by Crit. 4.	0.56
Crit. 3. by Crit. 5.	0.22
Crit. 4. by Crit. 5.	0.67
Total for Interactions	4.40
Total for All Effects	87.97
Residual	12.03
Total	100.00

Table 32

Comparison of Objective and Subjective Weights of Criteria: Judge Eight.

Criterion	Objective Weight	Subjective Weight	Absolute Difference
1	23	25	2
2	32	35	3
3	19	20	1
4	10	10	0
5	15	10	5
Mean Absolute Difference			2

Judge Eight subjectively identified no configural usage of pairs of criteria. Since no configural usage of pairs of cues was detected by objective analysis, Hypothesis 4.--that a judge will not correctly identify the configural use of pairs of cues detected by objective analysis--was rejected.

Judge Eight subjectively identified the use of an ordinal, but not interval, stimulus scale for all criteria. Since objective analysis detected the use of ordinal, but not interval, stimulus scales for all criteria, Hypothesis 5.--that a judge will not subjectively identify the type of stimulus scale determined by objective analysis--was rejected for all criteria.

FINDINGS FOR JUDGE NINE

As can be seen from an inspection of the ANOVA summary table (Table 33.), all main effects were significant (alpha=.05), as were the following interactions: Criterion 1. with Criterion 2.; Criterion 1. with Criterion 3.; Criterion 1. with Criterion 4.; Criterion 2. with Criterion 5.; and Criterion 3. with Criterion 4.

An ANOVA Summary table for the trend components of significant experimental effects is shown in Table 34. For the main effects, all linear trend components were significant (alpha=.05), and all quadratic trend components

Table 33

ANOVA Summary Table for Judge Nine: Strength of Proposal by Criteria.

Source of Variation	Sum of Square	DF	Mean Square	F	Signif. of F
Main Effects:					
Criterion 1	24.099	2	12.049	51.373	0.001
Criterion 2	4.173	2	2.086	8.896	0.001
Criterion 3	46.395	2	23.197	98.904	0.001
Criterion 4	37.802	2	18.901	80.586	0.001
Criterion 5	15.877	2	7.938	33.845	0.001
Interactions:					
Crit 1 by Crit 2	2.864	4	0.716	3.053	0.032
Crit 1 by Crit 3	2.864	4	0.716	3.053	0.032
Crit 1 by Crit 4	2.568	4	0.642	2.737	0.047
Crit 1 by Crit 5	0.716	4	0.179	0.763	0.557
Crit 2 by Crit 3	0.568	4	0.142	0.605	0.662
Crit 2 by Crit 4	0.938	4	0.235	1.000	0.423
Crit 2 by Crit 5	3.753	4	0.938	4.000	0.010
Crit 3 by Crit 4	2.938	4	0.735	3.132	0.029
Crit 3 by Crit 5	0.642	4	0.160	0.684	0.608
Crit 4 by Crit 5	1.012	4	0.253	1.079	0.384
Explained	147.210	50	2.944		
Residual	7.036	30	0.235		
Total	154.246	80	1.928		

Table 34

ANOVA Summary Table for Judge Nine: Trend Components for Significant Effects.

Source of Variation	Sum of Square	DF	Mean Square	F	Signif. of F
Criterion 1.					
Linear	24.000	1	24.000	102.331	0.000
Quadratic	0.099	1	0.099	0.422	0.521
Criterion 2.					
Linear	4.167	1	4.167	17.766	0.000
Quadratic	0.006	1	0.006	0.027	0.871
Criterion 3.					
Bilinear	46.296	1	46.296	197.398	0.000
Residual	0.099	1	0.099	0.421	0.521
Criterion 4.					
Linear	37.500	1	37.500	159.892	0.000
Quadratic	0.302	1	0.302	1.288	0.265
Criterion 5.					
Linear	15.574	1	15.574	66.405	0.000
Quadratic	0.303	1	0.303	1.292	0.265
Crit 1. by Crit 2.					
Bilinear	1.000	1	1.000	4.264	0.048
Residual	1.864	3	0.621	2.649	0.067
Crit 1. by Crit 3.					
Bilinear	2.778	1	2.778	11.844	0.002
Residual	0.086	3	0.029	0.123	0.946
Crit 1. by Crit 4.					
Bilinear	1.000	1	1.000	4.264	0.048
Residual	1.568	3	0.523	2.229	0.105
Crit 2. by Crit 5.					
Bilinear	0.000	1	0.000	0.000	1.000
Residual	3.753	3	1.251	5.334	0.005
Crit 3. by Crit 4.					
Bilinear	1.000	1	1.000	4.264	0.048
Residual	1.938	3	0.646	2.754	0.060
Residual	7.036	30	0.235		

were not significant (alpha=.05). Thus, Judge Nine used an interval stimulus scale for all criteria. The bilinear trend component was significant (alpha=.05), and the residual not significant (alpha=.05), for the following interactions: Criterion 1. with Criterion 2.; Criterion 1. with Criterion 3.; Criterion 1. with Criterion 4.; and Criterion 3. with Criterion 4. The bilinear trend component was not significant (alpha=.05) and the residual was significant (alpha=.05) for the interaction of Criterion 2. with Criterion 5. As a consequence of the occurrence of a significant residual, for Criterion 2. with Criterion 5., Hypothesis 2.--that a judge uses a linear model when integrating five simultaneously presented items of information--was rejected.

the $\underline{w}^2$ coefficients of the experimental effects for Judge Nine are displayed in Table 35. The following percentages of the total variance were accounted for by: main effects, 80.64%; interactions, 7.21%; all effects, 87.84%; residual, 12.16%. Since the percentage of the total variance accounted for by the experimentnal effects, 87.84%, exceeded 80%, Hypothesis 1.--that the judgement process can be quantified--was not rejected.

Data for comparing, for Judge Nine, the objective weights, $\underline{w}^2$, and subjective weights of criteria, are shown in Table 36. These data include the relative weights of

Table 35

Relative Magnitudes of Experimental Effects: Judge Nine.

Experimental Effect	$\underline{w}^2$
Main Effects:	
Criterion 1.	15.12
Criterion 2.	2.37
Criterion 3.	29.39
Criterion 4.	23.89
Criterion 5.	9.86
Total for Main Effects	80.64
Interactions:	
Crit. 1. by Crit. 2.	1.23
Crit. 1. by Crit. 3.	1.23
Crit. 1. by Crit. 4.	1.04
Crit. 1. by Crit. 5.	0.14
Crit. 2. by Crit. 3.	0.24
Crit. 2. by Crit. 4.	0.00
Crit. 2. by Crit. 5.	1.80
Crit. 3. by Crit. 4.	1.28
Crit. 3. by Crit. 5.	0.19
Crit. 4. by Crit. 5.	0.05
Total for Interactions	7.21
Total for All Effects	87.84
Residual	12.16
Total	100.00

Table 36

Comparison of Objective and Subjective Weights of Criteria: Judge Nine.

Criterion	Objective Weight	Subjective Weight	Absolute Difference
1	19	23	4
2	3	10	7
3	36	30	6
4	30	22	8
5	12	15	3
Mean Absolute Difference			6

criteria and the absolute differences between their objective and subjective weights. The mean absolute difference was 6 percentage points. The largest absolute difference, 8, was found for Criterion 4., and the smallest, 3, for Criterion 5. Since the mean absolute difference, 6, was less than 10 percentage points, Hypothesis 3.--that a judge's subjective weighting of criteria will differ from the weighting determined objectively--was rejected.

Judge Nine subjectively identified configural usage for the following pairs of criteria: Criterion 3. with Criterion 4., and Criterion 3. with Criterion 5. Since configural usage of pairs of criteria was objectively identified for Criterion 2. and Criterion 5., Hypothesis 4.--that a judge will not correctly identify the configural use of pairs of cues detected by objective analysis--was not rejected.

Judge Nine subjectively identified the use of an interval stimulus scale for all criteria. Since objective analysis detected the use of an interval stimulus scale for all criteria, Hypothesis 5.--that a judge will not subjectively identify the type of stimulus scale detected by objective analysis--was rejected for all criteria.

FINDINGS FOR JUDGE TEN

As can be seen from an inspection of the ANOVA summary table (Table 37.), all main effects were significant (alpha=.05), and all interactions were not significant (alpha=.05). As a consequence of the absence of significant interactions, Hypothesis 2.--that a judge uses a linear model when integrating five simultaneously presented cues--was not rejected.

An ANOVA summary table for the trend components of significant experimental effects is shown in Table 38. The linear trend components of all the main effects were significant (alpha=.05) while their quadratic trend components were not. Thus, interval stimulus scales were used for all criteria.

The $\underline{w}^2$ coefficients of the experimental effects for Judge Ten are displayed in Table 39. The following percentages of the total variance were accounted for by: main effects, 90.82%, interactions, 2.21%; all effects, 93.03%; residual, 6.97%. Since the percentage of the total variance accounted for by all the experimental effects, 93.03%, exceeded 80%, Hypothesis 1.--that the judgement process can be quantified--was not rejected.

The data for comparing, for Judge Ten, the objective weights, $\underline{w}^2$, and the subjective weights of criteria, are

Table 37

ANOVA Summary Table for Judge Ten: Strength of Proposal by Criteria.

Source of Variation	Sum of Square	DF	Mean Square	F	Signif. of F
Main Effects:					
Criterion 1	39.284	2	19.642	81.190	0.001
Criterion 2	62.691	2	31.346	129.567	0.001
Criterion 3	46.395	2	23.198	95.886	0.001
Criterion 4	78.247	2	39.123	161.716	0.001
Criterion 5	31.284	2	15.642	64.656	0.001
Interactions:					
Crit 1 by Crit 2	0.864	4	0.216	0.893	0.480
Crit 1 by Crit 3	1.160	4	0.290	1.199	0.332
Crit 1 by Crit 4	0.198	4	0.049	0.204	0.934
Crit 1 by Crit 5	1.383	4	0.346	1.429	0.249
Crit 2 by Crit 3	0.198	4	0.049	0.204	0.934
Crit 2 by Crit 4	2.123	4	0.531	2.194	0.094
Crit 2 by Crit 5	0.198	4	0.049	0.204	0.934
Crit 3 by Crit 4	1.975	4	0.494	2.041	0.114
Crit 3 by Crit 5	0.494	4	0.123	0.510	0.729
Crit 4 by Crit 5	1.531	4	0.383	1.582	0.205
Explained	268.024	50	5.360		
Residual	7.258	30	0.242		
Total	275.282	80	3.441		

Table 38

ANOVA Summary Table for Judge Ten: Trend Components for Significant Effects.

Source of Variation	Sum of Square	DF	Mean Square	F	Signif. of F
Criterion 1.					
Linear	39.185	1	39.185	161.967	0.000
Quadratic	0.099	1	0.099	0.408	0.528
Criterion 2.					
Linear	62.296	1	62.296	257.493	0.000
Quadratic	0.395	1	0.395	1.631	0.211
Criterion 3.					
Bilinear	46.296	1	46.296	191.360	0.000
Residual	0.099	1	0.099	0.408	0.528
Criterion 4.					
Linear	78.241	1	78.241	323.398	0.000
Quadratic	0.006	1	0.006	0.026	0.873
Criterion 5.					
Linear	31.130	1	31.130	128.670	0.000
Quadratic	0.154	1	0.154	0.638	0.431
Residual	7.258	30	0.242		

Table 39

Relative Magnitudes of Experimental Effects: Judge Ten.

Experimental Effect	w^2
Main Effects:	
Criterion 1.	13.79
Criterion 2.	22.11
Criterion 3.	16.32
Criterion 4.	27.64
Criterion 5.	10.95
Total for Main Effects	90.82
Interactions:	
Crit. 1. by Crit. 2.	0.04
Crit. 1. by Crit. 3.	0.07
Crit. 1. by Crit. 4.	0.27
Crit. 1. by Crit. 5.	0.15
Crit. 2. by Crit. 3.	0.27
Crit. 2. by Crit. 4.	0.41
Crit. 2. by Crit. 5.	0.27
Crit. 3. by Crit. 4.	0.36
Crit. 3. by Crit. 5.	0.17
Crit. 4. by Crit. 5.	0.20
Total for Interactions	2.21
Total for All Effects	93.03
Residual	6.97
Total	100.00

displayed in Table 40. These data include the relative weights of criteria and the absolute differences between their objective and subjective weights. The mean absolute difference was 4 percentage points. The largest absolute difference, 5, was found for Criterion 1. and Criterion 4., and the smallest, 2, for Criterion 3. All of the absolute differences were less than or equal to 10 percentage points. Since the mean absolute difference, 4, did not exceed 10 percentage points, Hypothesis 3.--that a judge's subjective weighting of criteria will differ from the weighting determined objectively--was rejected.

Judge Ten reported that the importance of the other criteria, in determining his judgement, varied with the value of criterion 4. If Criterion 4. was "High" or "Low," the importance of the other criteria was reduced. Thus, Judge Ten identified no pairs of cues as used configurally. Since objective analysis detected no configural usage of pairs of cues, Hypothesis 4.--that a judge will not correctly identify the configural use of pairs of cues detected by objective analysis--was rejected.

Judge Ten reported the use of an interval stimulus scale for all criteria. Since objective analysis detected the use of an interval stimulus scale, for all criteria, Hypothesis 5.--that a judge will not subjectively identify the type of stimulus scale detected by objective analysis--

Table 40

Comparison of Objective and Subjective Weights of Criteria: Judge Ten.

Criterion	Objective Weight	Subjective Weight	Absolute Difference
1	15	20	5
2	24	20	6
3	18	20	2
4	30	25	5
5	12	15	3
Mean Absolute Difference			4

was rejected for all criteria.

FINDINGS RELATED TO VALIDITY

Data relevant to consideration of the external validity of the data, for a judge, were: the percentage of the total variance accounted for by the experimental effects; whether or not the experimental task was reasonable[1]; and the occurrence of unlikely combinations of stimulus scale values.

The percentage of the total variance accounted for by the experimental effects ranged from 68.05% to 93.03% (as reported earlier in this chapter).

All judges considered the experimental task proposed in the initial interview, to be reasonable. Seven of the ten judges reported this belief in the final interview. Judge Four reported, in the final interview, that the five criteria he had selected as the most important for evaluating program proposals, were, in retrospect, not so. Further, Judge Four did not recall the meaning of the mnemonic for Criterion 5., which he had assigned in the initial interview. Neither of these considerations was expected to have affected the generality of Judge Four's responses with respect to: behaviour towards the stimuli; or

---

[1]Defined in Validity Considerations, Chapter 4.

the dimension of response (the components of external validity of the data). Judge Six reported that some of the criteria, chosen in the initial interview, were more appropriate to the overall program approval process rather than to rating program proposals which had been approved in principle. Further, Judge Six believed that his responses would be found to be "inconsistent and completely irrational." Judge Eight reported inappropriateness of the stimulus and response scales resulting from the constraint that the proposals were to be regarded as having been approved in principle. He felt that the response scale should have been degrees of "Strong" rather than ranging from "Weak" to "Strong." Further, Judge Eight reported difficulty in assigning a meaning to the stimulus scale values "High," "Medium," and "Low."

Judges One, Four, Six, Nine, and Ten reported that there were no unlikely combinations of stimulus scale values. The remaining judges reported that, although some combinations of stimulus scale values stimulus scale values were unlikely to be found in real program proposals, such combinations did not interfere with their judgement processes

SUMMARY OF THE FINDINGS

The research hypotheses were:

1. The judgement process can be quantified.
2. A judge uses a linear model when integrating five simultaneously presented cues.
3. A judge's subjective weighting of criteria will differ from the weighting determined objectively.
4. A judge will not correctly identify the configural use of pairs of cues detected by objective analysis.
5. A judge will not subjectively identify the type of stimulus scale detected by objective analysis.

The following hypotheses were rejected for Judge One: Hypotheses 1., 4., and 5. (for all criteria). The remaining hypotheses, Hypothesis 2., and Hypothesis 3., were not rejected for Judge One.

The following hypotheses were rejected for Judge Two: Hypotheses 4., and 5. (for Criteria 2., 3., 4., and 5.). The remaining hypotheses--Hypotheses 1., 2., 3., and 5. (for Criterion 1.)--were not rejected for Judge Two.

The following hypotheses were rejected for Judge Three: Hypotheses 3., 4., and 5. (for all criteria). The remaining hypotheses--Hypotheses 1. and 2.--were not rejected for Judge Three.

The following hypotheses were rejected for Judge

Four: Hypotheses 3., 4., and 5. (for Criteria 1. through 4.). The remaining hypotheses--Hypotheses 1., 2., and 5. (for Criterion 5.)--were not rejected for Judge Four.

The following hypotheses were rejected for Judge Five: Hypotheses 3. and 4.. The remaining hypotheses--Hypotheses 1., 2., and 5. (for all criteria)--were not rejected for Judge Five.

Hypotheses 3. and 5. (for Criterion 5.) were rejected for Judge Six, all other hypotheses were not rejected.

Hypothesis 3. was rejected for Judge Seven. The remaining hypotheses were not rejected for Judge Seven.

Hypotheses 3., 4., and 5. (for all criteria) were rejected for Judge Eight. Hypotheses 1., and 2. were not rejected for Judge Eight.

Hypotheses 2., 3., and 5. (for all criteria) were rejected for Judge Nine. Hypotheses 1., and 4. were not rejected for Judge Nine.

The following hypotheses were not rejected for Judge Ten: Hypothesis 1. and Hypothesis 2. The remaining hypotheses--Hypotheses 3., 4., and 5. (for all criteria)--were rejected.

Chapter 6

CONCLUSIONS, IMPLICATIONS AND SUGGESTIONS FOR FURTHER RESEARCH

This chapter contains: a restatement of the hypotheses; a summary of the results of hypothesis testing; a discussion of the validity of the data; a statement of the conclusions supported by the results of hypothesis testing, for each judge; a discussion of the implications of the findings; and suggestions for further research.

The research problem was stated in the fcrm of five research hypotheses:

1. The judgement process can be quantified.
2. A judge uses a linear model when integrating five simultaneously presented cues.
3. A judge's subjective weighting of criteria will differ from the weighting determined objectively.
4. A judge will not correctly identify the configural use of pairs of cues detected by objective analysis.
5. A judge will not subjectively identify the type of stimulus scale detected by objective analysis.

The results of testing these hypotheses are summarized in Table 41.

Table 41

Summary of the Results of Hypothesis Testing.

Hypothesis	Judge 1	2	3	4	5	6	7	8	9	10
1	R	N	N	N	N	N	N	N	N	N
2	N	N	N	N	N	N	N	N	R	N
3	N	N	R	R	R	R	R	R	R	R
4	R	R	R	R	R	N	N	R	N	R
5	R	R	R	R	N	N	N	R	R	R
Number[1]	5	4	5	4	5	4	5	5	5	5

R=rejected N=not rejected

[1]Number of criteria to which the decision applies for Hypothesis 5.

VALIDITY OF THE DATA

The percentage of the total variance accounted for by the experimental effects was an indicator of the generality, for a judge, of behaviour towards the stimuli, and the dimension of response. This indicator did not include behaviour in reponse to combinations of three or more cues. The experimental effects accounted for more than 80% of the total variance for all judges except Judge One. Judge One reported response to a combination of three cues which, if true, would account for the relatively low value, 68.05%, for that judge. Thus, on the basis of the percentage

of the total variance accounted for, that data were accepted as externally valid for each judge. That is, the data had generality for each judge with respect to: behaviour towards the stimuli; and the dimension of response.

The external validity of the data was confirmed for seven judges--Judges One, Two, Three, Five, Seven, Nine, and Ten--who considered (both before and after completing the experimental task) the experimental task to be reasonable. The concerns, relevant to the experimental task, of Judges Four, Six, and Eight, were discounted in view of the validity indicated by the percentage of the total variance accounted for by the experimental effects (87.61%, 80.65%, and 87.07%, respectively).

The absence of combinations of stimulus scale values which interfered with the judgement process, confirmed the external validity of the data, for each judge.

CONCLUSIONS

The conclusions supported by the results of testing the research hypotheses are presented for each judge.

Conclusions for Judge One

The rejection of Hypothesis 1. supported the conclusion that the judgement process, used by Judge One, was not quantifiable. The failure to reject Hypothesis 2. supported the conclusion that Judge One used a linear

model when integrating the five simultaneously presented cues. The failure to reject Hypothesis 3. supported the conclusion that Judge One did not understand the relative importance of the criteria determining his judgement. The rejection of Hypothesis 4. supported the conclusion that Judge One understood the way in which relationships among cues influenced his judgement. The rejection of Hypothesis 5. (for all criteria) supported the conclusion that Judge One understood the nature of the stimulus scale used.

Conclusions for Judge Two

The failure to reject Hypothesis 1. supported the conclusion that the judgement process, used by Judge Two, was quantifiable. The failure to reject Hypothesis 2. supported the conclusion that Judge Two used a linear model when integrating five simultaneously presented cues. The failure to reject Hypothesis 3. supported the conclusion that Judge Two did not understand the relative importance of the criteria in determining his judgement. The rejection of Hypothesis 4. supported the conclusion that Judge Two understood the way in which relationships among cues influenced his judgement. The rejection of Hypothesis 5., for Criteria 2. through 5., and the failure to reject that the hypothesis for Criterion 1., supported the conclusion that Judge Two did not understand the type of stimulus scale used for all criteria.

Conclusions for Judge Three

The failure to reject Hypothesis 1. supported the conclusion that the judgement process, used by Judge Three, was quantifiable. The failure to reject Hypothesis 2. supported the conclusion that Judge Three used a linear model when integrating five simultaneously presented cues. The rejection of Hypothesis 3. supported the conclusion that Judge Three understood the relative importance of the criteria in determining his judgement. The rejection of Hypothesis 4. supported the conclusion that Judge Three understood the way in which relationships ampng cues influenced his judgement. The rejection of Hypothesis 5. (for all criteria) supported the conclusion that Judge Three understood the nature of the stimulus scale used for all criteria.

Conclusions for Judge Four

The failure to reject Hypothesis 1. supported the conclusion that the judgement process, used by Judge Four, was quantifiable. The failure to reject Hypothesis 2. supported the conclusion that Judge Four used a linear model when integrating five simultaneously presented cues. The rejection of Hypothesis 3. supported the conclusion that Judge Four understood the relative importance of criteria in determining his judgement. The rejection of Hypothesis 4. supported the conclusion that Judge Four understood the

way in which relationships among cues influenced his judgement. The failure to reject Hypothesis 5., for Criterion 5., supported the conclusion that Judge Four did not understand the type of stimulus scale used for all criteria.

Conclusions for Judge Five

The failure to reject Hypothesis 1. supported the conclusion that the judgement process, used by Judge Five, was quantifiable. The failure to reject Hypothesis 2. supported the conclusion that Judge Five used a linear model when integrating five simultaneously presented cues. The rejection of Hypothesis 3. supported the conclusion that Judge Five understood the relative importance of criteria in determining his judgement. The rejection of Hypothesis 4. supported the conclusion that Judge Five understood the way in which relationships among cues influenced his judgement. The failure to reject Hypothesis 5., for any criterion, supported the conclusion that Judge Five did not understand the type of stimulus scale used for all criteria.

Conclusions for Judge Six

The failure to reject Hypothesis 1. supported the conclusion that the judgement process, used by Judge Six, was quantifiable. The failure to reject Hypothesis 2. supported the conclusion that Judge Six used a linear model when integrating five simultaneously presented cues.

The rejection of Hypothesis 3. supported the conclusion that Judge Six understood the relative importance of the criteria in determining his judgement. The failure to reject Hypothesis 4. supported the conclusion that Judge Six did not understand the way in which relationships among cues influenced his judgement. The rejection of Hypothesis 5., for Criterion 5., and the failure to reject that hypothesis for Criteria 1. through 4., supported the conclusion that Judge Six did not understand the nature of the stimulus scale used for all cues.

Conclusions for Judge Seven

The failure to reject Hypothesis 1. supported the conclusion that the judgement process, for Judge Seven, was quantifiable. The failure to reject Hypothesis 2. supported the conclusion that Judge Seven used a linear model when integrating five simultaneously presented cues. The rejection of Hypothesis 3. supported the conclusion that Judge Seven understood the relative importance of the criteria in determining his judgement. The failure to reject Hypothesis 4. supported the conclusion that Judge Seven did not understand the way in which relationships among cues influenced his judgement. The failure to reject Hypothesis 5., for any criterion, supported the conclusion that Judge Seven did not understand the nature of the stimulus scale used for all cues.

Conclusions for Judge Eight

The failure to reject Hypothesis 1. supported the conclusion that the judgement process, of Judge Eight, was quantifiable. The failure to reject Hypothesis 2. supported the conclusion that Judge Eight used a linear model when integrating five simultaneously presented cues. The rejection of Hypothesis 3. supported the conclusion that Judge Eight understood the relative importance of the criteria in determining his judgement. The rejection of Hypothesis 4. supported the conclusion that Judge Eight understood the way in which relationships among cues influenced his judgement. The rejection of Hypothesis 5. (for all criteria) supported the conclusion that Judge Eight understood the nature of the stimulus scale used for all criteria.

Conclusions for Judge Nine

The failure to reject Hypothesis 1. supported the conclusion that the judgement process, for Judge Nine, was quantifiable. The rejection of Hypothesis 2. supported the conclusion that Judge Nine did not use a linear model when integrating five simultaneously presented cues. The rejection of Hypothesis 3. supported the conclusion that Judge Nine understood the relative importance of the criteria in determining his judgement. The failure to reject Hypothesis 4. supported the conclusion that Judge Eight did

not understand the way relationships among cues influenced hsi judgement. The rejection of Hypothesis 5., for all criteria, supported the conclusion that Judge Nine understood the type of stimulus scale used for all criteria.

Conclusions for Judge Ten

The failure to reject Hypothesis 1. supported the conclusion that the judgement process, for Judge Ten, was quantifiable. The failure to reject Hypothesis 2. supported the conclusion that Judge Ten used a linear model when integrating five simultaneously presented cues. The rejection of Hypothesis 3. supported the conclusion that Judge Ten understood the relative importance of the criteria in determining his judgement. The rejection of Hypothesis 4. supported the conclusion that Judge Ten understood the way relationships among cues influenced his judgement. The rejection of Hypothesis 5., for all criteria, supported the conclusion that Judge Nine understood the nature of the stimulus scale, for all criteria.

Summary of Conclusions

The conclusion, that the judgement process can be quantified, was supported for all judges except Judge One. All judges, except Judge Nine, used a linear model when integrating five simultaneously presented cues. The following judges understood the relative importance of criteria in determining their judgement: Judge Three, Judge

Four, Judge Five, Judge Six, Judge Seven, Judge Eight, Judge Nine, and Judge Ten. The other judges--Judge One, and Judge Two--did not understand the relative importance of the criteria in determining their judgement. The following judges understood the way relationships among cues influenced their judgement: Judge One, Judge Two, Judge Three, Judge Five, Judge Eight, and Judge Ten. The other judges--Judge Four, Judge Six, Judge Seven, and Judge Nine--did not understand the way in which relationships among cues influenced their judgement. The following judges understood the nature of the stimulus scale used for all cues: Judge One, Judge Three, Judge Eight, Judge Nine, and Judge Ten. The other judges--Judge Two, Judge Four, Judge Five, Judge Six, and Judge Seven--did not understand, for all cues, the nature of the stimulus scale used.

IMPLICATIONS

The purpose of the study was to provide information about the judgement processes of selected judges in the context of a particular judgement task. The results of the study were not generalizable beyond the sample. The value of the results is their contribution to various aspects of the body of knowledge about human judgement, which has yet to be integrated into the mainstream of cognitive psychology. The conclusion, that the judgement process can be quantified,

was supported for nine of the ten judges studied. This confirms the findings of previous studies of human judgement. That this conclusion did not hold, for Judge One, may have been due to the presence of significant configural usage of three cues, as subjectively identified by that judge. Tests for the configural use of three or more cues were not allowed, by the fractional replicated design of this study, since third-and higher-order interactions were confounded.

Nine of the ten judges used a linear model when integrating information. This finding adds to previous studies, which have found support for the use of a linear model in judgement, support from the study of a task not previously investigated.

Only two of the ten judges did not understand how the relative importance of the criteria influenced their judgement. This does not agree with the findings of other studies. However, had a more stringent decision criterion been used, say an average difference of 2 instead of 10 percentage points, eight of the ten judges would have been said not to have understood how the relative importance of the criteria influenced their judgement.

Only four of the ten judges did not understand how configural usage of pairs of cues determined their judgement. This finding also does not agree with those of

previous studies. However, it is worth noting that although eight of the ten judges reported configural usage of more than two cues, the residuals containing the variance due to such usage were relatively small--less than 20% of the total variance--except for Judge One for whom the residual was 31.95%. This finding suggests that even if configural usage of more than two cues should prove statistically significant, such usage would only contribute slightly to the predictability of a judge's responses, which is consistent with previous findings.

Only three of the judges failed to identify the type of stimulus scales used for four, or more, of the criteria. Another two judges failed to identify the type of stimulus scale used for one criterion. The remaining six judges correctly identified the type of stimulus scales used for all criteria. This suggests that judges can distinguish between the use of interval and ordinal stimulus scales, when given a three-point--High, Medium, Low--scale.

In discussion, during the final interview, all judges expressed considerable interest in the results of the study. Judges exhibited most interest in results related to: the predictability of their judgements; evidence for their use of a simple linear model in integrating information; the relative importance of the criteria, as determined objectively; and the extent of the agreement or disagreement

between their objective and subjective weighting of criteria. This interest was due, in part, to a forthcoming self-evaluation of the procedure for reviewing program proposals. The results of the study provided information, about the nature and consistency of judgement for each judge, which has potential utility for improving: the consistency of judgement, for particular judges; and the compatibility of judgements across judges.

Much of the decision making activity of administrators involves the integration of information. Thus, information on the nature of the integration processes is relevant to the preparation of administrators. In addition to the presentation of theoretical aspects of information integration, the results of practical exercises, similar to the present study, would provide students with insights into their own cognitive processes.

The results of the study suggest that judges do not completely understand their own cognitive processes when combining five items of information. However, their judgements appear to be quantifiable. Thess factors suggest that studies designed to explicate the judgement process, followed by in-service training, might lead to greater consistency of judgement, both for individuals and groups.

The existence of inconsistencies in the combining of information, together with the ability of a linear model to

account for judgement, suggests that automated procedures might be used to improve the consistency of the judgement process. Such consistency could be expected to occur due to the assignment of the integration stage to a machine, or a predetermined algorithmic procedure.

SUGGESTIONS FOR FURTHER RESEARCH

The replication of the present study, with other samples, could be expected to provide data which, in conjunction with the data from the present study, would allow findings to be generalized to the population of interest, namely, judges involved in the evaluation of proposals for new instructional programs in higher education. Such replications would also provide data for cross-validating the experimental design.

Some judges subjectively identified configural usage of three or more cues. Confounding of third- and higher-order interactions, in the design of the study, did not allow tests for the presence of configural usage involving three or more cues. Thus, a study designed to detect such usage would be of value.

In attempting to identify the configural use of cues, some judges made comments which suggested that they used a summative rather than an averaging form of the linear model. The distinction between the two forms of the linear

model was not of interest in the present study. However, confirmation of the use of a linear model, for all but one of the judges studied, suggests that a study designed to distinguish between the two forms of the linear model, would be valuable.

Chapter 7

SUMMARY

This chapter contains summaries of: the statement of the problem; the research design; the methodology; the findings; and the conclusions of the study.

The Hypotheses

The purpose of the study was to investigate the way in which selected judges integrated information in a particular judgement task. The research problem was stated in the form of five hypotheses:

1. The judgement process can be quantified.
2. A judge uses a linear model when integrating five cues which are presented simultaneously.
3. A judge's subjective weighting of criteria will differ from the weighting determined objectively.
4. A judge will not correctly identify the configural use of pairs of cues detected by objective analysis.
5. A judge will not subjectively identify the type of stimulus scale detected by objective analysis.

The ANOVA Paradigm

A one-third fractional replication of a 3^5 factorial design was used to construct the stimuli. Consequently third- and higher-order interactions were confounded. The factors were five criteria for evaluating proposals for new instructional programs in higher education. Each factor had three levels: High, Medium, and Low. Response was required on a nine-point numeric scale with the verbal anchors "weak," "average," and "strong."

Classical ANOVA was used to test for the presence of significant experimental effects. Further, classical ANOVA provided the data for computing the relative importance, $\underline{w}^2$, of the experimental effects. The trend components of significant experimental effects were then tested for significance. A significant quadratic trend component of a main effect was interpreted as indicative of the use of an ordinal, but not interval, stimulus scale. A significant residual for an interaction, after partitioning out the bilinear trend component, was interpreted as indicative of the configural usage of a pair of criteria (response to patterns of those cues).

Each stimulus was defined by five cues. Each cue consisted of a criterion for evaluating proposals and an associated rating on a three-point scale. The context of the judgement task was defined in relation to the actual task of

evaluating proposals for new instructional programs. Judges were asked to consider the proposals, represented by the stimuli, to have been approved in principle, and thus to have met or exceeded some minimum standard on each criterion. Further, judges were asked to consider each stimulus as representing a program proposal which had been rated, by themselves or others, on each criterion, independently of the other criteria. The purposes of these constraints were: to minimize the conjunctive use of cues; and to establish a link between the simulated stimuli and real proposals.

A Judge's Objective Policy

For the purposes of this study, a judge's *objective policy* was defined as the way in which he integrated information, as described through the application of the ANOVA paradigm, in making a judgement. Elements of a judge's objective policy, of interest in this study, were: the extent to which the judgement process could be quantified; the goodness of fit of a linear model of judgement; the non-linear use of cues; and the relative weights placed on cues or patterns of cues.

A Judge's Subjective Policy

A judge's *subjective policy* was defined as the way in which he said that he integrated information in making a judgement. The subjective policy of each judge was

represented by three measures; an estimate of the relative importance (weights) of the five criteria in determining the response; the identification, by the judge, of configural usage of cues; and the identification, by the judge, of the stimulus scale as interval or only ordinal.

Summary of the Findings

The following hypotheses were rejected for Judge One: Hypotheses 1., 4., and 5. (for all criteria). The remaining hypotheses, Hypothesis 2., and Hypothesis 3. , were not rejected for Judge One.

The following hypotheses were rejected for Judge Two: Hypotheses 4., and 5. (for Criteria 2., 3., 4., and 5.). The remaining hypotheses--Hypotheses 1., 2., 3., and 5. (for Criterion 1.)--were not rejected for Judge Two.

The following hypotheses were rejected for Judge Three: Hypotheses 3., 4., and 5. (for all criteria). The remaining hypotheses--Hypotheses 1. and 2.--were not rejected for Judge Three.

The following hypotheses were rejected for Judge Four: Hypotheses 3., 4., and 5. (for Criteria 1. through 4.). The remaining hypotheses--Hypotheses 1., 2., and 5. (for Criterion 5.)--were not rejected for Judge Four.

The following hypotheses were rejected for Judge Five: Hypotheses 3. and 4.. The remaining hypotheses--Hypotheses 1., 2., and 5. (for all criteria)--were not

rejected for Judge Five.

Hypotheses 3. and 5. (for Criterion 5.) was rejected for Judge Six, all other hypotheses were not rejected.

Hypothesis 3. was rejected for Judge Seven. The remaining hypotheses were not rejected for Judge Seven.

Hypotheses 3., 4., and 5. (for all criteria) were rejected for Judge Eight. Hypotheses 1. and 2. were not rejected for Judge Eight.

Hypotheses 2., 3., and 5. (for all criteria) were rejected for Judge Nine. Hypothesis 1., and 4. were not rejected for Judge Nine.

The following hypotheses were not rejected for Judge Ten: Hypothesis 1. and Hypothesis 2. The remaining hypotheses--Hypotheses 3., 4., and 5. (for all criteria)--were rejected.

Summary of Conclusions

The conclusion, that the judgement process can be quantified, was supported for all judges except Judge One. All judges, except Judge Nine, used a linear model when integrating five simultaneously presented cues. The following judges understood the relative importance of criteria in determining their judgement: Judge Three, Judge Four, Judge Five, Judge Six, Judge Seven, Judge Nine, and Judge Ten. The other judges--Judge One, Judge Two, and Judge Eight--did not understand the relative importance of the

criteria in determining their judgement. The following judges understood the way relationships among cues influenced their judgement: Judge One, Judge Two, Judge Three, Judge Five, Judge Eight, and Judge Ten. The other judges--Judge Four, Judge Six, Judge Seven, and Judge Nine--did not understand the way in which relationships among cues influenced their judgement. The following judges understood the nature of the stimulus scale used for all cues: Judge One, Judge Three, Judge Eight, Judge Nine, and Judge Ten. The other judges--Judge Two, Judge Four, Judge Five, Judge Six, and Judge Seven--did not understand, for all cues, the nature of the stimulus scale used.

BIBLIOGRAPHY

Abelson, R.P. et. al. (eds)
1968 Theories of Cognitive Consistency: A Sourcebook, Chicago: Rand McNally and Co.

Anderson, N.H.
1967 "Application of a Weighted Averaging Model to a Psychophysical Averaging Task," Psychon. Sci., 8(6):227-28.

Anderson, N.H.
1968 "A Simple Model for Integrating Information," in Abelson, 1968:731-41.

Anderson, N.H.
1969 "Comment on 'An Analysis-of-variance model for the Assessment of Configural Cue Utilization in Clinical Judgement'," Psych. Bull., 72(1):63-65.

Anderson, N.H.
1970 "Functional Measurement and Psychophysical Judgement," Psych. Rev., 77(3):153-70.

Houston, S.R.
1974 Judgement Analysis: Tool for Decision Makers, New York: MSS Information Corporation.

Knox, R.F. and Hoffman, P.J.
1962 "Effects of Variation of Profile Format on Intelligence and Sociability Judgements," J. App. Psych., 46:14-20.

Runkel, P.J. and McGrath, J.E.
1972 Research on Human Behaviour, U.S.A.: Holt, Rinehart and Winston Inc.

Slovic, P.
1969 "Analyzing the Expert Judge: A Descriptive Study of a Stockbroker's Decision Processes," J. App. Psych., 53(4):255-63.

Slovic, P. and Lichtenstein, S.
1971 "Comparison of Bayesian and Regression Approaches to the Study of Information Processing and Judgement," Organizational Behaviour and Human Performance, 6:649-744.

Winer, B.J.
1971 *Statistical Principles in Experimental Design*, New York: McGraw-Hill Book Company.

APPENDIX 1

INSTRUCTIONS FOR RESPONDING TO STIMULI AND SAMPLE STIMULUS.

INSTRUCTIONS

The following program proposal profiles represent proposals which have been:

1. APPROVED IN PRINCIPLE, and
2. rated independently on each of the five criteria.

You should evaluate the strength of each proposal given the information in its profile, and circle the appropriate number on the 9-point scale provided.

Please mail the completed instrument to me or phone 432-3094 and I will collect it.

GOOD LUCK!

Tony Marshall,
Dept. of Edcuational Administration,
The University of Alberta,
Edmonton, Alberta.

SAMPLE PROGRAM PROPOSAL 1 10

Criterion	Rating
1. Approriateness of social and/or economic objectives.	HIGH
2. ratio of costs to benefits.	LOW
3. relevance to current priorities.	HIGH
4. justification of the need for a new program.	MEDIUM
5. justification of the program elements.	MEDIUM

Strength of Proposal:

weak			average			strong		
1	2	3	4	5	6	7	8	9

CIRCLE ONE

APPENDIX 2

MNEMONICS FOR CRITERIA SELECTED BY JUDGES DURING THE INITIAL INTERVIEWS

Table 42

Mnemonics for Criteria: Judge One.

Criterion	Mnemonic
1.	evidence for appropriateness of philosophy and purpose.
2.	justification of implementation and maintenance costs.
3.	suitability of evaluation procedures.
4.	evidence for suitability of the curriculum.
5.	evidence for compatibility with existing priorities.

Table 43

Mnemonics for Criteria: Judge Two.

Criterion	Mnemonic
1.	appropriateness of philosophy and objectives.
2.	internal consistency and suitability of program-curriculum design.
3.	approriateness of program to institution.
4.	medium and long-range appropriateness (financial-economic).
5.	effectiveness of mechanisms for evaluating and changing the program.

Table 44

Mnemonics for Criteria: Judge Three.

Criterion	Mnemonic
1.	need identification.
2.	conceptual relevance.
3.	operational feasibility.
4.	identification of change indicators.
5.	potential effects.

Table 45

Mnemonics for Criteria: Judge Four.

Criterion	Mnemonic
1.	employment opportunities on graduation.
2.	adequacy of basic life skills component.
3.	extent to which the program meets the needs of special groups.
4.	appropriateness of skills to intended job.
5.	approriateness of institutional vs. on-the-job training.

Table 46

Mnemonics for Criteria: Judge Five.

Criterion	Mnemonic
1.	feasibility of goals.
2.	evidence for employment demand.
3.	justification of capital and operating costs.
4.	evidence for suitability of curriculum.
5.	evidence for maximization of use of existing resources.

Table 47

Mnemonics for Criteria: Judge Six.

Criterion	Mnemonic
1.	justification of need (demand for graduates).
2	evidence for student demand.
3.	justification of operating costs.
4.	evidence of evaluation procedures.
5.	consistency with institutional purpose.

Table 48

Mnemonics for Criteria: Judge Seven.

Criterion	Mnemonic
1.	documentation of need for program.
2.	evidence for support from reference groups.
3.	evidence for maximization of use of existing learning experiences and facilities.
4.	evidence of appropriateness of institutional basis for program.
5.	extent of integration with existing programs.

Table 49

Mnemonics for Criteria: Judge Eight.

Criterion	Mnemonic
1.	evidence fpr appropriateness of philosophy and purpose.
2.	justification of competency basis of curriculum.
3.	evidence for availability of employment opportunities for graduates.
4.	justification of implementation costa.
5.	evidence for satisfactory evaluation procedure.

Table 50

Mnemonics for Criteria: Judge Nine.

Criterion	Mnemonic
1.	philosophy, purpose, and performance-referenced goals.
2.	justification of implementation and maintenance costs.
3.	justification of demand and monitoring procedures.
4.	relevance of curriculum to goals and performance criteria.
5.	evidence for compatability with existing arrangements.

Table 51

Mnemonics for Criteria: Judge Ten.

Criterion	Mnemonic
1.	appropriateness of social and/or economic objectives.
2.	ratio of costs to benefits.
3.	relevance to current priorities.
4.	justification of the need for a new program.
5.	justification of the program elements.